



UNIVERSIDADE FEDERAL DO OESTE DO PARÁ
PROGRAMA DE PÓS-GRADUAÇÃO DOUTORADO SOCIEDADE, NATUREZA E DESENVOLVIMENTO

**REFERENCIAL SEMÂNTICO NO SUPORTE DA IDENTIFICAÇÃO BOTÂNICA
DE ESPÉCIES AMAZÔNICAS**

MÁRCIO JOSÉ MOUTINHO DA PONTE

Santarém – Pará
Abril/2017

MÁRCIO JOSÉ MOUTINHO DA PONTE

**REFERENCIAL SEMÂNTICO NO SUPORTE DA IDENTIFICAÇÃO BOTÂNICA
DE ESPÉCIES AMAZÔNICAS**

Orientador

Prof. Dr. Celson Pantoja Lima

Co-orientador

Prof. Dr. Lauro Euclides Soares Barata

Tese submetida ao Programa de Pós-Graduação em Sociedade Natureza e Desenvolvimento da Universidade Federal do Oeste do Pará como requisito obrigatório para obtenção do grau de Doutor Interdisciplinar, área de concentração: Gestão do Conhecimento e Inovação para o Desenvolvimento Sustentável.

Santarém - Pará

Mar/2017

O tempo não para! Só a saudade é que faz as coisas pararem no tempo ...”
Mário Quintana

Ao meu irmão **Marcelo Moutinho da Ponte** (*in memoriam*), por toda a amizade.

À minha família, Nicole, minha esposa, e Luísa e Samuel, meus filhos, pelo amor, paciência e tolerância das minhas ausências.

Aos meus pais, Joviniano Ferreira da Ponte e Isabel do Carmo Moutinho da Ponte, que foram os primeiros incentivadores e me ensinaram o valor do conhecimento.

AGRADECIMENTOS

A quem mais posso agradecer, antes de tudo, senão ao nosso Pai? Muito obrigado pelas oportunidades oferecidas, assim como pelas adversidades impostas. Sem vossos fardos, dificilmente teria ombros fortalecidos para prosseguir pelo percurso traçado.

À minha esposa, Nicole, por toda dedicação, amor e companheirismo dedicado neste difícil trajeto, fundamental para a conclusão deste trabalho.

Aos filhos, Luísa e Samuel, pela tempestade de alegria e felicidade emanada todos os dias, e por me fazer lembrar a todo o momento da importância da vida, e que tenho que fazer valer a pena cada instante quando estou em suas companhias.

Ao querido irmão Marcelo (in memoriam), por todo o amor, amizade e cumplicidade dedicado, pelo exemplo de força e superação em meio a tantas adversidades.

Agradeço também aos meus bons pais, Joviniano e Isabel, que para fornecer o melhor que podiam a mim e a meus irmãos, muitas vezes abdicavam tanto de si. Muito obrigado!

Ao prof. Celson Lima, orientador e amigo, que contribuiu para o desenvolvimento do trabalho e para minha formação como pessoa e pesquisador, obrigado pelo apoio e dedicação.

Ao prof. Ricardo Jardim-Gonçalves, orientador da co-tutela, obrigado pela confiança depositada e pelas valorosas experiências de trabalho na Universidade Nova de Lisboa.

Ao primo Victor Moutinho, pela amizade, incentivo e dedicação na coleta dos materiais.

Ao Co-orientador Lauro Barata pelas valiosas contribuições.

Aos gajos amigos de Portugal, Ricardo Jardim-Gonçalves, Paulo Figueiras, Ruben Costa, Pedro Oliveira e Raquel pela amizade, incentivo e atenção no decorrer do período de co-tutela.

Agradeço ao PPGSND/UFOPA, PPGEEC/UNL, a UNINOVA e ao grupo de pesquisa GRIS por oportunizar a conclusão deste trabalho.

Aos membros da Banca Examinadora desta tese (qualificação e defesa final), pelas valorosas contribuições.

Agradeço também as demais pessoas que participaram direta ou indiretamente do desenvolvimento deste trabalho, em especial aos familiares, aos amigos de longa data e aos amigos de trabalho.

Muito obrigado!

SUMÁRIO

SUMÁRIO.....	5
RESUMO	8
ABSTRACT	9
LISTA DE FIGURAS	10
LISTA DE TABELAS	12
LISTA DE ACRÔNIMOS	13
1. INTRODUÇÃO	1
1.1 Objetivo e Resultados Esperados.....	2
1.2 Motivação e Caracterização do Problema	3
1.3 Pergunta da Pesquisa e Hipótese	4
1.4 Contribuição Inédita	5
1.5 Contextualização do Trabalho	6
1.6 Estrutura da Tese	7
2. FUNDAMENTAÇÃO TEÓRICA	9
2.1 Gestão do Conhecimento	9
2.1.1 O Conhecimento.....	9
2.1.2 Modelo de GC.....	14
2.1.4 Referencial Semântico: A Ontologia.....	27
2.1.5 Vetor Semântico.....	30
2.2 Os Sistemas Especialistas (SEs)	34
2.2.1 Classificação dos Sistemas Especialistas	36
2.2.2 Conhecimento Especialista.....	36
2.2.3 Arquitetura de um SE	41
2.3 Reconhecimento de Padrões	44
2.3.1 Reconhecimento de padrões de imagem.....	46
2.3.2 Segmentação: Histograma	48

2.3.3 Extração de Características: Descritores de Haralick	50
2.3.4 Classificação: Redes Neurais Artificiais	52
3. TRABALHOS CORRELATOS	57
3.1 Reconhecimento de Padrão de Imagem Botânica.....	57
3.1.1 Captura da Amostra	59
3.1.2 Segmentação	61
3.1.3 Classificação	65
3.1.4 Taxas de Reconhecimento de Padrão	67
3.2 Ontologia Ambiental.....	68
3.2.1 Manejo Florestal.....	69
3.2.2 Monitoramento e Impacto Ambiental	70
3.2.3 Ecossistema e Biodiversidade	71
3.2.4 Tecnologias Associadas	72
3.3 Sistemas especialistas e chaveamentos no processo de identificação botânica.....	74
3.4 Análise Comparativa.....	74
4. METODOLOGIA DA PESQUISA	76
4.1 Método clássico	76
4.2 Metodologias Complementares	77
4.2.1 Metodologia para o desenvolvimento da Ontologia	77
4.2.2 Metodologia para o desenvolvimento dos Sistemas Especialistas	81
4.3 Procedimentos Metodológicos.....	84
4.3.1 Ambiente de Pesquisa: Amazônia - Tapajós.....	84
4.3.2 Levantamento das principais espécies comercializadas na região amazônica	86
4.3.3 Caracterização das amostras de madeira	86
5. REFERENCIAL SEMÂNTICO APLICADO AO PROCESSO DE IDENTIFICAÇÃO BOTÂNICA	88
5.1 Definições e conceitos utilizados neste trabalho	88
5.2 Modelo conceitual da tese.....	90
5.2.1 Visão	90
5.2.2 Instanciação do Modelo	91
5.2 O Projeto	92
5.2.1 Processo de criação do Referencial Semântico	93
5.2.2 Especificação de Requisitos.....	94
5.2.2.1 Visão Funcional.....	95

5.2.2.2 Visão Arquitetural	98
5.2.2.3 Visão Comportamental.....	99
5.3 Implementação	100
5.2.1 Tecnologias usadas na implementação.....	101
5.2.2 A Ontologia ONTO-AMAZONTIMBER	102
5.2.3 Sistema Especialista para identificação botânica – Inventário Florestal.....	107
5.2.3 Sistema Especialista para classificação de imagem de madeira	117
5.4 Análise dos Resultados & Validação	133
5.4.1 Restrições e delimitação do escopo	133
5.4.2 Validação Ontologia ONTO-AMAZONTIMBER.....	133
5.4.2 Análise dos resultados do Sistema Especialista para identificação botânica – Inventário Florestal	134
6. CONCLUSÕES E TRABALHOS FUTUROS	139
6.1 Visão Geral da Tese	139
6.2 Contribuições da Tese.....	141
6.3 Desafios Encontrados	145
6.4 Trabalhos Futuros	146
7. REFERÊNCIAS BIBLIOGRÁFICAS	148
8. APÊNDICES	172
8.1 Cooperação Internacional	172
8.1.1 Base de conhecimento para aquicultura.....	173
8.1.2 Sistema especialista para diagnóstico de doenças de peixes	173
8.1.3 Sistema especialista para investigação de empresas de aquicultura	175

RESUMO

A identificação botânica de espécies vegetais nativas da Amazônia é parte integrante do inventário florestal, imprescindível para o plano de manejo florestal e essencial para que a comunidade científica conheça mais e melhor a floresta Amazônica. No entanto, o processo usual de identificação botânica normalmente usa apenas o conhecimento empírico de nativos conhecedores da floresta (mateiros), os quais adotam nomes vernaculares (populares) na determinação das espécies, que por sua vez, apresentam divergências dos nomes científicos catalogados por taxonomistas. Tendo esta problemática como cenário de pesquisa, este trabalho propõe um modelo conceitual para suportar um referencial semântico que apoie o processo de identificação de espécies botânicas da Amazônia, com intuito de minimizar as divergências de conhecimento entre taxonomistas e mateiros, e consequentemente aumentar a acurácia do método de identificação. Para tal, são utilizados recursos semânticos (e.g. ontologia e vetores semânticos) na formalização do conhecimento capturado. Dois cenários de aplicação são usados para avaliar este trabalho, nomeadamente: (i) o cenário Inventário Florestal que utiliza como instrumento avaliativo o sistema especialista para identificação botânica por características; (ii) o cenário Imagem Madeira que utiliza como instrumento avaliativo o sistema especialista para classificação de imagem de madeira. Como parte dos resultados, estes cenários utilizam o reconhecimento de padrão no apoio à tomada de decisão usando ferramentas computacionais no auxílio ao processo de identificação de espécies florestais comercializadas na Amazônia, com taxas de acertos de 65% de reconhecimento em imagens de madeira. Por conseguinte conclui-se que o referencial semântico proposto neste trabalho contribui sobremaneira no âmbito ambiental, no que tange à produção de conhecimento sobre a Amazônia.

Palavras-chaves: Gestão do Conhecimento, reconhecimento de padrão, ontologia, vetor semântico, sistemas especialistas.

ABSTRACT

The botanical identification of species native of the Amazonian is an integral part of any forest inventory, mandatory for forest management plan and, therefore, essential for the scientific community to know more and better the Amazonian forest. However, the usual process of botanical identification is based on the empirical knowledge of native people from the forest (Bushmen) that use vernacular names to identify species, which in turn do not match the scientific names cataloged by taxonomists. Having this problem as the research scenario, this work proposes a conceptual model based on a semantic referential to support the process of identification of botanical species in the Amazon, helping reducing the knowledge mismatch between taxonomists and foresters, and consequently increase the accuracy of the current identification method. Semantic resources (e. g. ontology and semantic vector) are used in the formalization of captured knowledge. Two application scenarios are used to assess this work, namely: (i) the Forest Inventory that uses as an evaluation tool the specialist system for botanical identification by characteristics; (ii) the Image Timber that uses as an evaluation tool the expert system for image classification of wood. As part of the results, these scenarios use the pattern recognition to support decision making by providing computational tools to aid the process of identification of forest species marketed in the Amazon, with 65% accuracy rates in wood images. Final conclusion is that the semantic reference proposed in this work brings a relevant contribution regarding the production of knowledge about the Amazon area.

Keywords: Knowledge Management, pattern recognition, ontology, semantic vector, expert systems

LISTA DE FIGURAS

Figura 1. Estrutura da formação do conhecimento	13
Figura 2. Espiral SECI ou processo SECI	16
Figura 3. Espiral do conhecimento	18
Figura 4. Modelo de GC de Prax	19
Figura 5. Modelo de GC de Cavalcanti, Gomes e Neto.	20
Figura 6. Modelo de Angeloni	20
Figura 7. Modelo de GC de Stollenwerk.....	21
Figura 8. O ciclo do conhecimento. Fonte: Choo (2006).....	22
Figura 9. Modelo de Wiig (WIIG, 1993).....	23
Figura 10. Modelo ICAS (Bennet e Bennet, 2004).....	26
Figura 11. Distância euclidiana entre dois vetores	32
Figura 12. Arquitetura de um SE	42
Figura 13. As três fases do reconhecimento de padrões de imagens digitais	47
Figura 14. À esquerda uma imagem I, ao centro o histograma da imagem em valores ($h(I)$) e em porcentagem ($p(I)$), à direita uma representação do histograma de forma gráfica.....	49
Figura 15. Topo à esquerda, imagem em nível de cinza, e abaixo dela o histograma da imagem. Topo à direita, a mesma imagem após equalização, e o histograma equalizado da imagem.	50
Figura 16. Linearidade das redes neurais (KROGH e VEDELSBY, 1995)	55
Figura 17. Fase de Propagação (LNCC 2008).....	55
Figura 18. Fase de Retropropagação (LNCC 2008).....	56
Figura 19. Processo de reconhecimento de padrão.....	57
Figura 20. Imagens Madeira.....	58
Figura 21. Etapas do método científico.....	77
Figura 22. Metodologia e-COGNOS para construção da Ontologia	79
Figura 23. Ferramenta Protégé	81
Figura 24. Metodologia do desenvolvimento dos sistemas especialistas.	82
Figura 25. FLONA Tapajós, adaptado (ICMBio, 2017)	85
Figura 26. Lixadeira de cinta portátil.....	87
Figura 27. Microscópio digital	87
Figura 28. O Modelo Conceitual.....	91
Figura 29. Instanciação do Modelo Conceitual	92
Figura 30. Visão geral do Projeto	93
Figura 31. Processo de desenvolvimento da estrutura semântica	94
Figura 32. Diagrama de Caso de Uso - Atribuições dos atores	95
Figura 33. Diagrama de Caso de Uso – Gestão do conhecimento	97
Figura 34. . Diagrama de caso do uso – Interface	97
Figura 35. Diagrama de Componentes.....	98
Figura 36. Diagrama de Sequência – Inventário florestal	99
Figura 37. Diagrama de Sequência – Madeira	100
Figura 38. Tecnologias utilizadas.....	101
Figura 39. Base de conhecimento do domínio.....	103

Figura 40. Estrutura taxonômica da ontologia.....	104
Figura 41. Ontologia – Label inglês/português	104
Figura 42. Instanciação do Modelo Conceitual	106
Figura 43. Disjoint Class.....	107
Figura 44. Relações semânticas obtidas pelo método <i>getObjectsFromObjectTriple</i>	110
Figura 45. Generalização de resultados	112
Figura 46. Especificação de resultados	113
Figura 47. Representação do Vetor Semântico.....	114
Figura 48. Exemplo de Vetor Semântico	115
Figura 49. Arquitetura do SE	117
Figura 50. Matriz de coocorrência	119
Figura 51. Respectivamente: Maçaranduba(M01):0.35515; Tauari (T06):0.3506.....	122
Figura 52. As três fases do reconhecimento de padrões de imagens digitais	128
Figura 53. Arquitetura da RNA desenvolvida	129
Figura 54. Configuração da RNA.....	130
Figura 55. Gráfico de regressão linear para lixas 40, 80 e 120 respectivamente.	130
Figura 56. Taxa de reconhecimento	131
Figura 57. Imagens macroscópicas de madeiras amazônicas, dispostas por aplicação de lixa: (a) Bagassa guianensis (tatajuba); (b) Carapa guianensis (andiroba); (c) Cedrelinga cateniformis (cedrorana); (d) Dipteryx ferrea (cumarú); (e) Goupia glabra (cupiúba); (f) Handroanthus sp. (Ipê); (g) Hymenolobium petraeum (angelim); (h) Manilkara spp. (massaranduba); (i) Peltogeny spp. (Pau-roxo); (j) Vataireopsis sp. (fava).	132
Figura 58. Validação Ontologia Onto-AmazonTimber	134
Figura 59. Interface – Sistema Especialista	135
Figura 60 Ontologia AquaSmart	173
Figura 61 SE para diagnóstico de doenças	175
Figura 62 Sistema especialista para investigação de empresas de aquicultura - Características.....	176
Figura 63 Sistema especialista para investigação de empresas de aquicultura - Nome.....	177

LISTA DE TABELAS

Tabela 1. Frases e definições associadas a Conhecimento	9
Tabela 2. Definição de Padrão.....	45
Tabela 3. Reconhecimento de Padrões (RP)	46
Tabela 4. Imagens.....	59
Tabela 5. Descritores de textura	61
Tabela 6. Classificadores	65
Tabela 7. Taxa de reconhecimento	67
Tabela 8. Aplicação da Ontologia	68
Tabela 9. Algumas tecnologias usadas.....	72
Tabela 10. Métodos List() da API JENA.....	109
Tabela 11. Características de Haralick.....	119
Tabela 12. Variável tipo da tabela 11.....	120
Tabela 13. Agrupamento com características selecionadas	121
Tabela 14. A característica Soma das Médias para o conjunto de Imagens.	122
Tabela 15. Agrupamento com K-means para Tauari e Maçaranduba (primeira execução)	123
Tabela 16. Agrupamento com K-means para Tauarí e Maçaranduba (segunda execução)	124
Tabela 17. Agrupamento com K-means para Tauari e Maçaranduba (com duas variáveis)	125
Tabela 18. Classificação por RNA para Tauarí e Maçaranduba	126
Tabela 19. Classificação por RNA para Tauarí e Maçaranduba (com duas variáveis).....	127
Tabela 20. Publicações	142
Tabela 21. Trabalhos acadêmicos	144

LISTA DE ACRÔNIMOS

ACV – Avaliação do Ciclo de Vida

FCT – Faculdade de Ciência e Tecnologia

GC – Gestão do Conhecimento

GRIS – Grupo de Investigação sobre Interoperabilidade de Sistemas

IBAMA – Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis

ITTO – *International Tropical Timber Organization*

KNN – (do inglês. K- nearest neighbors)

LN – Linguagens Naturais

LPF – Laboratório de Produtos Florestais

VSM - Modelo de Espaço Vetorial

MLP - *Multilayer Perceptron*

RNA – Redes Neurais Artificiais

RP – Reconhecimento de Padrões

SBC – Sistemas Baseados em Conhecimento

SFB – Serviço Florestal Brasileiro

SVM – (do inglês. *Support Vector Machine*)

UFOPA – Universidade Federal do Oeste do Pará

UNINOVA – Instituto para o Desenvolvimento de Novas Tecnologias

UNL – Universidade Nova de Lisboa

1. INTRODUÇÃO

A Gestão do Conhecimento¹ (GC) está intimamente relacionada ao fator sucesso no processo de tomada de decisões, o que tende a aumentar à medida que se intensifica a interação entre a produção de conhecimento e tecnologia (ROSSETTI e MORALES, 2007). Nesta perspectiva, o atual contexto tecnológico apresenta novos desafios aos processos de produção de conhecimento, o que permite expandir o alcance nos domínios de aplicação e potencializar as funcionalidades de um referencial semântico.

Segundo Luckesi e Passos (1996), adquirir conhecimento não é compreender a realidade retendo informação, mas utilizando-se desta para desvendar o novo e avançar, porque quanto mais competente for o entendimento do mundo, mais satisfatória será a ação do sujeito que o detém. Desta forma, o conhecimento é o resultado do processamento analítico de informações que fornece subsídios essenciais para tomada de decisões estratégicas. Em outras palavras, as informações facilitam no processo de cognição, mas, por si só, não realizam efetivamente o conhecimento.

Le Coadic (2004) vislumbrou duas características que marcam a atualidade do acesso à informação: sua explosão quantitativa (crescimento exponencial e facilidade de publicação de conteúdos por parte dos usuários) e a implosão do tempo para a sua comunicação (o limite humano de absorção de informações e interação é limitado). Segundo este autor, *“a conjunção desses dois fenômenos levou ao aparecimento de fluxos [...] de informação muito elevados, isto é, a circulação de consideráveis quantidades de informação por unidade de tempo”* (LE COADIC, 2004).

O uso de referenciais semânticos (como, por exemplo, ontologias² e taxonomias de conceitos³) para representar conhecimento na web e em outros domínios permitiu o desenvolvimento de novos mecanismos de buscas (e.g. buscas semânticas), inferências e análise em banco de dados e conteúdos complexos, mapas conceituais, além de viabilizar o

¹ Gestão do Conhecimento: possui a finalidade de entender, focar e gerir, de forma sistemática, bem como explicitar para deliberar sobre a construção de conhecimento, sua reutilização, renovação e aplicação. (WIIG, 1997)

² Ontologia: é uma especificação de conhecimento consensual sobre um modelo abstrato de domínio, definida explicitamente em termos de conceitos, suas propriedades e relações por meio de axiomas, possibilitando, assim, que seja legível por máquinas (STUDER et al., 1998).

³ Taxonomia de conceitos: representar conceitos através de termos, agilizar a comunicação entre especialistas e outros públicos; encontrar o consenso; propor formas de controle da diversidade de significação e oferecer um mapa de área que servirá como guia em processo de conhecimento (TERRA, 2000).

uso de agentes inteligentes para buscar informação na Web de forma muito mais fidedigna e rápida (BITTENCOURT et al., 2008). Contudo, o desenvolvimento de um referencial semântico é um processo complexo, demorado, visto que se trata de um artefato da engenharia de software com vocabulários, ferramentas e metodologias específicas que envolvem profissionais especialistas no domínio de conhecimento a ser aplicado e engenheiros do conhecimento (GUARINO, 1998).

O presente trabalho propõe um modelo conceitual que engloba a construção de um referencial semântico aplicado ao domínio da Botânica, capaz de suportar os processos de reconhecimento de padrão aquando da identificação de espécies da flora amazônica, o que implica em necessariamente imergir no cenário botânico e angariar o conhecimento com profissionais e fontes de conhecimentos disponíveis. Parte da pesquisa, mais especificamente a integração do referencial semântico com as tecnologias de reconhecimento de padrões, foi desenvolvida no Grupo de Investigação sobre Interoperabilidade de Sistemas (GRIS) integrado no Instituto UNINOVA - Instituto para o Desenvolvimento de Novas Tecnologias, sediado na Faculdade de Ciências e Tecnologia (FCT) da Universidade Nova de Lisboa (UNL), Portugal.

1.1 Objetivo e Resultados Esperados

O objetivo geral deste trabalho é propor um referencial semântico que apoie o processo de identificação botânica de espécies amazônicas.

Os resultados esperados para este trabalhos de tese são apresentados em quatro grupos, nomeadamente resultados científicos, tecnológicos, académicos e outros.

RESULTADOS CIENTÍFICOS

- Um modelo conceitual que aporte um referencial semântico no âmbito da botânica, para suportar processos de reconhecimento de padrões.
- Um modelo conceitual de suporte ao processo de integração semântica entre o referencial semântico e o processo de reconhecimento de padrões.

RESULTADOS TECNOLÓGICOS

- Provas de conceito baseadas em ambientes computacionais que ajudem a validar os conceitos propostos na tese.
- Ferramentas computacionais que viabilizem a tomada de decisão no processo de identificação botânica.

- Algoritmos de reconhecimento de padrões de imagens e conhecimento de espécies amazônicas.

RESULTADOS ACADÊMICOS

- “Integração de descritores de textura (Haralick e coloração) para reconhecimento de padrão de espécies florestais da Amazônia” Trabalho de Conclusão de Curso do Curso Ciência da Computação – UFOPA, em desenvolvimento.
- “Madeiras agrupadas comercialmente sob o mesmo nome vernacular no Oeste do Pará” Iniciação Científica - UFOPA

OUTROS RESULTADOS BASEADOS NOS CENÁRIOS DE AVALIAÇÃO DA TESE

- Contribuição no projeto AquaSmart, implementação das tecnologias e conceitos utilizados na tese.
- Potenciais sugestões para o aprimoramento do desenvolvimento sustentável no manejo florestal no setor madeireiro.
- Potenciais sugestões para a contribuição na conservação ambiental e o desenvolvimento social na região Amazônica.
- Potenciais sugestões para uma melhor fiscalização no setor madeireiro.

1.2 Motivação e Caracterização do Problema

A constante preocupação pela conservação da Biodiversidade tornou-se elemento integrante da consciência da comunidade internacional, motivada pela percepção de uma série de ameaças, como a destruição crescente e contínua da vegetação nativa, visando à implantação de culturas agropastoris ou à extração madeireira, sem a manutenção das áreas de reserva legal e proteção permanente. Problema igualmente sério é o desmatamento, com posterior abandono do solo, deixando-o descoberto, sujeito ao empobrecimento e à erosão.

A floresta Amazônica, detentora da maior diversidade biológica do planeta, tornou-se foco das atenções da opinião pública, uma vez que diversas espécies estão em extinção propiciando a diminuição dos recursos naturais em decorrência da degradação, em um ritmo sem precedente e colocando em risco o bem-estar da humanidade (BRADSHAW et al., 2009).

A fragmentação de informações disponíveis, assim como a grande variedade de fontes científicas e não científicas dificultam o entendimento do domínio de conhecimento como um todo e evidenciam a necessidade de integração de conhecimentos dispersos.

A busca por informações fidedignas no âmbito da botânica são agravadas pela escassez de botânicos taxonomistas que atuem na região, mais especificamente no processo de

identificação botânica de espécies Amazônicas, atividade esta prevista no inventário florestal obrigatoriamente contida em todo sistema de manejo florestal. Tal situação leva à substituição do profissional por nativos conhecedores da floresta (mateiros), os quais adotam nome vernacular (i.e, popular ou não científico) na determinação das espécies.

A inexistência de um conceito que considere conhecimento, padrões e os atores intervenientes neste cenário motiva a condução deste trabalho. A ausência de um referencial semântico com uma base de conhecimento sólida no âmbito da botânica dificulta interagir ou mesmo imergir com áreas de conhecimentos distintos, em destaque a área da computação e tecnologia como um todo, permitindo novos cenários, possibilidades e desafios na botânica.

Neste contexto, identifica-se a importância do domínio da botânica nos mais diversos setores do conhecimento e, desta forma, vislumbra-se a necessidade de construção de conhecimento necessário para criar um referencial semântico na área. Mais que isso, propor um conceito que permita que este processo de criação seja replicável para que possibilite a expansão do referencial semântico desenvolvido ou mesmo o desenvolvimento em outros domínios do conhecimento.

1.3 Pergunta da Pesquisa e Hipótese

A seguinte pergunta de pesquisa guia este trabalho:

Como representar formalmente o conhecimento necessário para criar um referencial semântico no domínio da botânica, suportando o reconhecimento de padrões de espécies amazônicas?

Esta pergunta de pesquisa sustenta-se nas seguintes hipóteses:

- *O uso de um referencial semântico no domínio da botânica diminui a inconsistências nos resultados do processo de identificação de espécies amazônicas;*
- *O uso de reconhecimento de padrão aumenta a acurácia do processo de identificação botânica;*
- *A integração entre o referencial semântico e as tecnologias de reconhecimento de padrões aumenta as taxas de precisão dos resultados de classificação de espécies botânicas.*

1.4 Contribuição Inédita

A definição desta pesquisa foi baseada em trabalhos disponíveis na literatura no âmbito da produção de conhecimento no domínio da identificação botânica que serão referenciados no decorrer da tese. Entretanto, nota-se a escassez de um referencial semântico que suporte a identificação botânica. Desta forma, a contribuição inédita desta tese é a criação de um referencial semântico que suporte o processo de identificação botânica de espécies amazônicas. Para tal, são considerados aspectos essenciais para construção do referencial semântico, que apresentam ineditismos e contribuições científicas decorrente da elaboração e da validação do modelo conceitual proposto, nomeadamente:

- Um modelo conceitual que aporte um referencial semântico no âmbito da botânica, suportando processos de reconhecimento de padrões;
- Um referencial semântico que aporte a produção de conhecimento no domínio da botânica mais especificamente no processo de identificação botânica;
- A ontologia Onto-AmazonTimber (representando conceitos em Português e em Inglês também) que oferece relações semânticas e restrições axiomáticas que inferem sobre características e imagens botânicas dispostas hierarquicamente, e representadas por entidades e propriedades;

No decorrer do desenvolvimento da tese destacam-se outras contribuições relevantes que apesar de não serem inéditas, contribuíram sobremaneira para o enriquecimento científico do trabalho, à citar:

- Um sistema especialista para identificar espécies botânicas baseados em reconhecimento de padrões de características e imagens das espécies;
- Um sistema especialista para classificar imagens de madeira de espécies amazônicas;
- Algoritmos de reconhecimento de padrões de imagens e conhecimento de espécies amazônicas
- Atribuição de vetor semântico como máquina de inferência para melhoria da classificação de padrões botânicos;
- A integração semântica entre o referencial semântico e o reconhecimento de padrões;

1.5 Contextualização do Trabalho

A representação formal de conhecimento com suas complexas estruturas de relações é o objetivo de um referencial semântico (LEGG, 2007). No presente trabalho, o referencial semântico será concretizado com o desenvolvimento de tecnologias semânticas, dentre estas destaca-se a ontologia. Na concepção de Breitman (2005):

Ontologias são especificações formais e explícitas de conceitualizações compartilhadas. Ontologias são modelos conceituais que capturam e explicitam o vocabulário utilizado nas aplicações semânticas. Servem como base para garantir uma comunicação livre de ambiguidades.

As ontologias identificam um núcleo de conceitos e suas relações no intuito de representar um domínio de conhecimento. Não limitando-se somente a isto, mas ao se fornecer uma descrição exata do conhecimento de forma diferente da linguagem natural, amplamente utilizada nas pesquisas sintáticas onde a recuperação da informação é limitada a buscas por palavra-chave, as ontologias possibilitam o “entendimento” pelos programas computacionais.

Por isso, há necessidade do desenvolvimento de ferramentas que ampliem a dimensão semântica dos termos para processamento pelo computador. As relações semânticas com propósito de representação do conhecimento, ao serem explicitadas, eliminam interpretações dúbias: dois conceitos podem se relacionar de muitas formas diferentes, mas ao explicitar a relação pretendida, o conhecimento é transmitido (CAMPOS, 2009).

Nesta tese, a ontologia representa uma parte do domínio de conhecimento da Botânica, considerando a complexidade do domínio como o vasto acervo da biodiversidade das espécies florestais existentes na maior floresta tropical do mundo (Amazônia), aliado a falta de acurácia do atual processo de identificação botânica e as adversidades da exploração madeireira na região.

O processo de comercialização da madeira inicia-se no inventário florestal, é a base do plano de manejo e da produção de uma empresa madeireira, um dos mais importantes passos rumo ao manejo sustentável. Este deve fornecer, além do volume total explorável, a distribuição do número de árvores por hectare e por classe de diâmetro; área basal por hectare, por classe de diâmetro, por grupo de espécies e para cada espécie individualmente (SILVA, 2001).

O método usual de inventário florestal segundo Procópio e Secco (2008) conta com o conhecimento empírico de nativos conhecedores da floresta, os quais adotam nomes

vernaculares na identificação das espécies. O nome científico correspondente, listado pelas empresas madeireiras no inventário comercial, geralmente vem de listas de nomes prováveis para o nome vernacular estabelecido pelo mateiro, proveniente de literatura especializada ou de listas do IBAMA. Procópio e Secco (2008) afirmam que a correspondência de nomes é feita sem a adoção de critérios científicos, morfológicos ou ecológicos das espécies.

A utilização de nomenclaturas populares nos inventários florestais torna inconsistente a verdadeira ocorrência geográfica das espécies, visto que a denominação vernacular varia conforme a região, a cultura ou ao uso na comercialização (MARTINS-DA-SILVA, 2002).

Não existe uma padronização da nomenclatura científica com a denominação popular, posto que uma espécie pode receber diversos nomes vernaculares ou várias espécies podem ser designadas por um único nome vernacular. Essa problemática encontra-se muito bem expressa no livro lançado pelo Laboratório de Produtos Florestais do Ibama (CAMARGOS et al., 1996).

A exemplo disto, Souza e colaboradores (2006) identificam espécies de gênero *Dipteryx* Schreber e *Tabebuia* Gomes ex DC sendo comercializadas em conjunto sob a denominação comum de Cumaru, salientando que este último é popularmente conhecido como Ipê. Outro fato interessante no estudo é a representatividade das mesmas, onde *Dipteryx odorata*, espécie habitualmente associada ao nome vernacular de Cumaru, ficou como a segunda mais representativa da amostragem.

Moutinho e equipe (2008), ao realizarem o levantamento das espécies comercializadas como pau-mulato no Estado do Amapá, identificaram seis espécies pertencentes a quatro gêneros botânicos, em que duas dessas espécies são comumente conhecidas como cumaru-amarelo (*Taralea oppositifolia*) e cupiúba (*Goupia glabra*). Sobre outras duas, pertencentes ao gênero *Capirona* Spruce, sequer havia registro de sua comercialização.

Neste contexto, busca-se nesta pesquisa integrar o referencial semântico da botânica com tecnologias de reconhecimento de padrão que possam apoiar a tomada de decisão no processo de identificação botânica, possibilitando o aumento da acurácia dos resultados.

1.6 Estrutura da Tese

Este documento está estruturado como se segue:

- Capítulo II, Fundamentação Teórica: é composto de um referencial teórico no qual se apresentam as áreas GC, Sistemas Especialistas e Reconhecimento de Padrões (Processamento de Imagem);
- Capítulo III, Trabalhos Correlatos: contém os trabalhos mais relevantes identificados na literatura que estão relacionados com a problemática discutida nesta tese;
- Capítulo IV, Metodologia da Pesquisa: apresenta os aspectos metodológicos da tese, o que inclui metodologias para construção da Ontologia, de aquisição de conhecimento e para validação do referencial semântico;
- Capítulo V, Modelo do Referencial Semântico aplicado ao processo de identificação botânica: apresenta os principais conceitos e definições que embasam esta tese, descreve o modelo conceitual da tese proposto por meio de uma descrição detalhada sobre as três etapas que o compõem, sendo: (1) Projeto; (2) Implementação; e (3) Análise de Resultados e Validação;
- Capítulo VI, Conclusões e Trabalhos Futuros: é dedicado às conclusões obtidas com esta tese, apresenta as contribuições da pesquisa e os desafios encontrados, além de sugerir possibilidades de trabalhos futuros a partir dos resultados obtidos com esta pesquisa.



2. FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta a fundamentação teórica que sustenta o desenvolvimento deste trabalho, cobrindo aos seguintes temas: (i) teorias, conceitos e ferramentas que suportam a GC, incluindo o conhecimento propriamente dito, e aspectos técnicos e estruturais dos Referenciais Semânticos, com destaque para as ontologias; e (ii) teorias e técnicas de Reconhecimento de Padrões baseado em Processamento de Imagem.

2.1 Gestão do Conhecimento

2.1.1 O Conhecimento

A perspectiva de um conceito que defina conhecimento não pode satisfazer unanimamente a comunidade científica, posto que o conhecimento não pertence a uma área específica da ciência; ele é universal e transita por caminhos diversos com o propósito de atender novos paradigmas.

O debate existente em torno do entendimento conceitual sobre o conhecimento perpassa gerações e períodos da humanidade. A Tabela 1, apoiada em Nonaka e Takeuchi (1997), apresenta visões do conhecimento cunhadas por diversos autores em diferentes momentos da humanidade, oferecendo elementos sobre a amplitude do conceito, sua complexidade e suas variações, que acompanham o movimento evolutivo da sociedade.

Tabela 1. Frases e definições associadas a Conhecimento
Fonte: Adaptado de (NONAKA e TAKEUCHI, 1997)

AUTOR	CITAÇÕES / DEFINIÇÃO
Aristóteles (384-322 a.C)	O conhecimento é sempre ocasionado pela percepção sensorial.
René Descartes (1596-1650)	A verdade definitiva é deduzida a partir de um “eu pensante”, independente do corpo e da matéria.
John Locke (1632-1704)	Compara a mente a uma tábula rasa, uma folha de papel sem conteúdo. Somente as experiências podem proporcionar ideias à mente, sendo possível adquirir conhecimento por indução, a partir de experiências sensoriais.

Immanuel Kant (1724-1804)	O conhecimento parte do pensamento lógico do racionalismo e da experiência sensorial do empirismo em que a mente humana é tábula rasa ativa, que ordena as experiências sensoriais no tempo e no espaço, suprindo-se de conceitos como ferramenta de compreensão.
Georg W. F. Hegel (1770-1831)	O conhecimento começa com a percepção sensorial, ao se tornar mais subjetiva e mais racional por meio da purificação dialética dos sentidos chega, por fim, ao estágio do conhecimento do espírito absoluto.
Jena-Paul Sartre (1905-1980)	O mundo se revela pela nossa conduta, é a escolha intencional do fim que revela a realidade.
Nonaka e Takeuchi (1997)	O conhecimento é percebido como um dos mais importantes ativos que uma empresa ou pessoas podem possuir, e que suplanta os tradicionais fatores de produção, terra, capital e trabalho.
Sveiby (1998)	A construção do conhecimento é contínua e cumulativa, em um processo de consumo e usos constantes, onde a informação armazenada não possui valor, mas sua preciosidade está inserida no contexto da geração do conhecimento.
Davenport e Prusak (1998)	É a informação nutrida de valor que interage com aspectos relativos à experiência, contexto, interpretação e reflexão.
Terra (2000)	Elemento reutilizável, cujo valor se amplia à medida que é utilizado.
Zabot (2002)	É um fator competitivo de extremo poder e importância, não só sua aquisição, como também sua criação e transferência.
Leff (2002)	É condicionado pelo contexto geográfico, ecológico e cultural em que se produz e se reproduz determinada formação social.
Servin (2005)	É derivado da informação, mas alerta que é o valor incorporado que dá o significado ao conhecimento. Nesse sentido, é possível afirmar que a informação desprovida de valor não gera conhecimento.
Figueiredo (2005)	É aquilo que torna alguém apto a agir em circunstâncias específicas.

Conhecimento é o objeto central da GC, discutida a seguir. Segundo Wiig (1997), a GC possui a finalidade de entender, focar e gerir, de forma sistemática, bem como explicitar para deliberar sobre a construção de conhecimento, sua reutilização, renovação e aplicação.

De acordo com Singh e Soltanai (2010), as mudanças no ambiente onde se interage aliado à complexidade tecnológica exigem uma estrita organização do conhecimento baseado em informações. A GC enquadra-se em situações críticas de adaptação a ambientes conturbados, remete aos recursos tecnológicos aliados à informação e à criatividade dos envolvidos nos momentos de recuperação.

Por sua vez, Amin, Zawawi e Timan (2011) relatam que a GC, do ponto de vista estratégico, desenvolve a aprendizagem organizacional, que vai muito além do aprendizado individual englobando também as relações estabelecidas entre estes indivíduos, nas organizações. O conjunto de diferentes habilidades empregadas em locais distintos (ao cooperarem para executar uma tarefa comum) cria uma coexistência que ultrapassa a própria tarefa para a criação de um novo conhecimento organizacional.

Ampliando as possibilidades oferecidas pela GC, Fialho e colaboradores (2006) argumentam que a evolução do pensamento humano tem sido um grande aliado no desenvolvimento da GC, juntamente com as tecnologias de informação e comunicação. A relação do homem com as tecnologias tem promovido a gestão de pessoas, papel fundamental no amparo à gestão de conhecimento, proporcionando um processo de mudança de cultura nas organizações.

Amin, Zawawi e Timan (2011) destacam que o conhecimento organizacional é oriundo do compartilhamento do conhecimento de cada um dos indivíduos que a organização compõe. Armazenar este conhecimento organizacional torna-se uma vantagem estratégica para, então, compartilhar e atender novas solicitações que necessitam dos conhecimentos; tal partilha de conhecimento é definida como “cultura de interação social”.

2.1.1.1 ESTRUTURA E FORMAÇÃO DO CONHECIMENTO

De acordo com Davenport e Prusak (1998, p. 1),

“por mais primário que possa soar, é importante frisar que dado, informação e conhecimento não são sinônimos [...] entender o que são esses três elementos e como passar de um para outro é essencial para a realização bem-sucedida do trabalho ligado ao conhecimento”.

Bhatt (2001) enfatiza a dificuldade de definir dado, informação e conhecimento. Apenas por meio da perspectiva de um usuário, é que se consegue distingui-los. Normalmente, dados são considerados fatos novos, enquanto a informação é vista como um conjunto de dados organizados. O conhecimento, por sua vez, é percebido como uma informação estruturada e com sentido claro.

Firestone e McElroy (2001) diferenciam dado, informação e conhecimento da seguinte maneira:

- Dado: é um valor observável, mensurável ou calculável de um atributo. O contexto (sempre existente) é o que torna compreensível a estrutura do formato de um dado.
- Informação: sempre promovida por dados. Em termos gerais, informação é definida como dado mais interpretações.
- Conhecimento: é a informação que passou por testes e avaliações em processos que procuram eliminar erros e alcançar a verdade; portanto, mais confiável e aprimorado por registros e experiências.

Desta forma, identifica-se a ocorrência de uma relação hierárquica de valores entre dados, que, indexados e contextualizados, geram a informação e, por conseguinte, quando analisados e integrados no processo cognitivo (humano) geram o conhecimento, como pode ser observado na Figura 1.

No entanto, uma hierarquia pode ser analisada no sentido inverso e, desta forma, a informação emerge com a existência de um conhecimento que possibilita compreender sua formulação e os dados são catalogados após a informação que valida a ocorrência de fatos (TUOMI, 1999). O conhecimento passa a existir na ocorrência da interpretação, análise e avaliação da informação dentro de um contexto e de um modelo mental. Como cada indivíduo possui um modelo mental único, específico e distinto, essas informações podem transformar-se em conhecimentos distintos (ALBINO; GARAVELLI; SHIUMA, 2001)(GUPTA; MCDANIEL, 2002)(ROWLEY, 2007)(BENNET; BENNET, 2008).

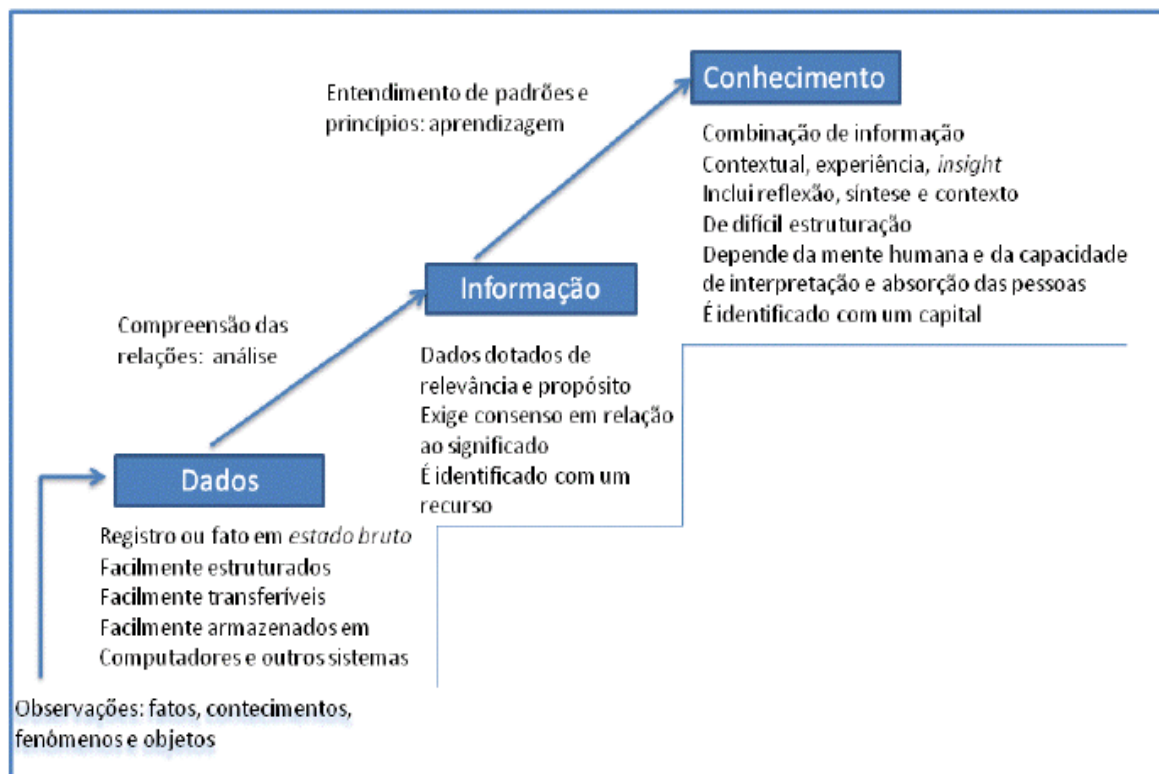


Figura 1. Estrutura da formação do conhecimento

Fonte: Adaptado de Beal (2004)

2.1.1.2 CONHECIMENTO TÁCITO X CONHECIMENTO EXPLÍCITO

Segundo Nonaka e Takeuchi (1997) a criação do conhecimento apresenta duas dimensões em sua estrutura conceitual básica de formulação: epistemológica e ontológica. A dimensão epistemológica retrata as distinções entre o conhecimento tácito e o explícito, por sua vez, a dimensão ontológica retrata os níveis de envolvimento das entidades no processo de criação do conhecimento, nomeadamente individual, grupal, organizacional e interorganizacional.

Na dimensão epistemológica, segundo Fialho e colaboradores (2006), o conhecimento tácito envolve fatores intangíveis, como as perspectivas e sistemas de valor do ser humano. Também possui uma importante dimensão cognitiva, com esquemas, modelos mentais, crenças e percepções. Por conseguinte, o conhecimento explícito pode ser transmitido facilmente entre os indivíduos, comunicado e compartilhado de maneira simples sob a forma de dados brutos, fórmulas científicas, procedimentos codificados ou princípios universais, podendo ser processado, armazenado e transmitido eletronicamente de forma rápida.

De acordo com Nonaka e Takeuchi (1997), o conhecimento tácito passa a ser adquirido pela memória de experiências de um indivíduo e com o surgimento de uma tarefa adversa novas experiências surgem. Em contraste, o conhecimento explícito compõe sua fração do conjunto do conhecimento formal e sistematizado, por meio de palestras, seminários, livros e vídeos, entre outras formas.

Embora os tipos de conhecimento *tácito* e *explícito* sejam distintos, na sua essência não são entidades totalmente separadas e podem ser entendidas como complementares. O conhecimento tácito é pessoal, específico ao contexto e, assim, difícil de ser formulado e comunicado. Já o conhecimento explícito (ou “codificado”) refere-se ao conhecimento transmissível em linguagem formal e sistemática (NONAKA e TAKEUCHI, 1997). Um espiral do conhecimento “surge quando a interação entre conhecimento tácito e conhecimento explícito eleva-se dinamicamente de um nível ontológico inferior até níveis mais altos” (NONAKA e TAKEUCHI, 1997).

2.1.2 Modelo de GC

Apesar de existir uma vasta bibliografia sobre a GC, encontram-se poucos modelos formais. Segundo Wilson (2002) modelos são essencialmente descrições de entidades, processos, atributos e suas interações no intuito de assertivas e viabilização. Neste sentido, faz-se necessário identificar os principais modelos e práticas que destacam-se na literatura, cujas experiências buscam reportar à criação de conhecimento.

Todos os modelos apresentados neste capítulo são relevantes, e cada um oferece fundamentos teóricos valiosos na GC e na compreensão das organizações da atualidade. Na sua totalidade compartilham uma abordagem conexcionista e holística para entender melhor a natureza do conhecimento como um sistema complexo adaptativo que inclui conhecedores, ambiente organizacional e sua rede de compartilhamento de conhecimento.

Segundo Dalkir (2005) os mais relevantes modelos de GC são: Von Krogh e Ross; Nonaka e Takeuchi; Prax; Cavacanti; Angeloni; Stollenwerk; Chun Wei Choo; Wiig; Boisoti I-Space. Tais modelos são descritos a seguir.

2.1.2.1 MODELO EPISTEMIOLÓGICO ORGANIZACIONAL VON KROGH E ROOS

O modelo de GC de Von Krogh e Roos (1995) apresenta uma distinção entre o conhecimento individual e o conhecimento social, aplicando uma abordagem epistemológica para a GC organizacional.

A perspectiva cognitivista (VARELA, 1992) propõe que um sistema cognitivo (cérebro humano ou computador) cria representações (isto é, modelos de realidade) e que o aprendizado ocorre quando estas representações são manipuladas. A epistemologia organizacional cognitiva entende o conhecimento organizacional como um sistema auto-organizado em que as informações dos seres humanos são transparentes para o exterior. O cérebro é uma máquina baseada na lógica e dedução que não permite quaisquer proposições contraditórias, comparada à organização que escolhe as informações de seu ambiente e as processa de forma lógica.

A abordagem conexionista, por outro lado, é mais abrangente. O cérebro percebe a "totalidade", propriedades globais, padrões e sinergias, não limitando-se a processos de sequenciamento de símbolos. As regras de aprendizagem ocorrem com a interação entre os diversos componentes dessas redes. As decisões não são apenas tomadas a partir do ambiente, mas também geradas internamente. Os indivíduos são partes de um sistema organizacional interligado, e o conhecimento, por sua vez, é um fenômeno emergente que se origina a partir das interações sociais destes indivíduos.

Nesta perspectiva, o conhecimento reside não só na mente dos indivíduos, mas também nas conexões entre esses indivíduos. A mente coletiva é formada como a representação desta rede, e é isso que está no cerne da GC organizacional.

Von Krogh e Roos (1995) adotam a abordagem conexionista em seu modelo de GC de epistemologia organizacional, o conhecimento reside nas relações entre os indivíduos dos mais diversos níveis sociais. Conhecimento é dito ser "encarnado", isto é, "tudo que se sabe é conhecido por alguém" (VON KROGH e ROOS, 1995).

2.1.2.2 MODELO DE NONAKA E TAKEUCHI

O modelo de Gestão de Conhecimento de Nonaka e Takeuchi, conhecido como modelo SECI, apresenta as seguintes dimensões especificadas no processo de criação do conhecimento por Nonaka e Takeuchi (1997): (i) a dimensão epistemológica, que evidencia a transferência do conhecimento estritamente por linguagem formal e sistemática de indivíduos, únicos detentores da capacidade de criar conhecimento; e (ii) a dimensão ontológica que relata a criação do conhecimento baseado na interação entre o conhecimento tácito e o explícito em ambiente de GC. Tais conhecimento tácito e explícito assumem um papel determinante na definição das fases de conversão de conhecimento (TAKEUCHI e NONAKA, 2008). Deste modo, o processo de conversão do conhecimento é composto pelas

fases **Socialização, Externalização, Combinação e Internalização** (SECI, a sigla que identifica o modelo), definidas como se segue:

- **Socialização:** consiste em compartilhar o conhecimento em interações face a face, naturais e tipicamente sociais.
- **Externalização:** apresenta uma forma visível ao conhecimento tácito e converte-o em conhecimento explícito. Pode ser definido como “um processo de criação de conhecimento, na medida em que o conhecimento tácito se torna explícito, tomando as formas de metáforas, analogias, conceitos, hipóteses ou modelos” (Nonaka e Takeuchi, 1995, p.4)
- **Combinação:** ocorre quando os conceitos são classificados e estruturados em um sistema de conhecimento.
- **Internalização:** é o processo de conversão ou integração de experiências e conhecimentos partilhados e / ou individuais em modelos mentais individuais.

A criação do conhecimento é um processo iterativo contínuo e dinâmico envolvendo o conhecimento tácito e o explícito. Esta interação ocorre com a transformação do conhecimento apresentada no Modelo SECI apresentado na Figura 2.

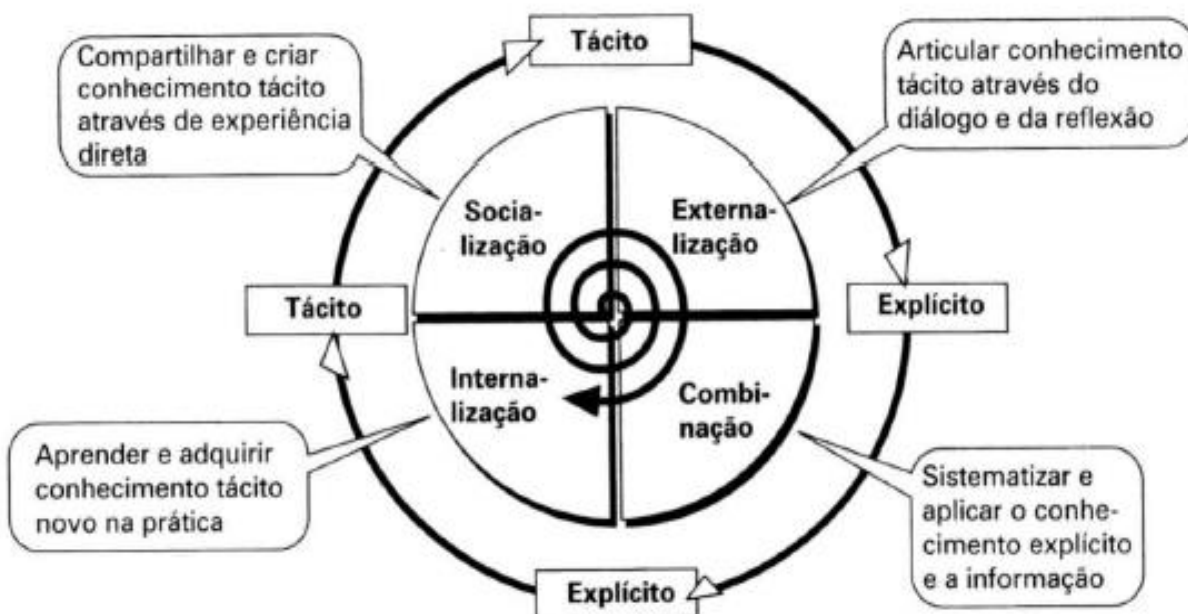


Figura 2. Espiral SECI ou processo SECI

Fonte: (Takeuchi e Nonaka, 2008).

A espiral do conhecimento se inicia na fase de Socialização com a construção de um campo de interação. Esse campo tem por função facilitar o compartilhamento das experiências e dos modelos mentais dos integrantes das organizações. A fase Externalização, por sua vez, é desencadeada pelo diálogo ou pela reflexão coletiva, para os quais o uso da metáfora ou da analogia auxilia os integrantes da equipe a articularem o conhecimento tácito que, de outra forma, seria difícil de comunicar. A Combinação, por sua vez, é desencadeada pela rede do conhecimento recentemente criada e do conhecimento existente de outros setores da organização, cristalizando-os em um novo produto ou sistema administrativo.

Neste contexto, observa-se a organização contemporânea como um grande sistema complexo que interage com o seu meio interno e externo, trocando sensações e experiências exatamente como acontece com os seres-vivos. Desta forma, entende-se organização como um processo que necessite da cooperação de pessoas para a resolução de um problema ou execução de uma tarefa em um determinado âmbito do conhecimento.

Segundo Nonaka e Takeuchi (1997), uma organização para se estabelecer nessa nova era tem que se adaptar aos mecanismos de criação do conhecimento e proporcionar condições para que as 4 transformações do conhecimento, anteriormente citadas, sejam plenamente praticadas.

Para os autores, não basta que as transformações aconteçam de forma isolada e única dentro da organização; é necessário fazer funcionar o “motor” socialização – exteriorização – combinação – interiorização, de forma que no final de um ciclo, o produto final seja um novo conhecimento a ser interiorizado, realinhando novamente o ciclo de forma repetitiva para assim conseguir construir uma “espiral virtuosa do conhecimento”.

A Figura 3 apresenta graficamente a espiral virtuosa do conhecimento descrito por Nonaka e Takeuchi (1997). Este espiral mostra que quanto mais o conhecimento tende a ser explícito – estratégico – organizacional, maior propulsão e alcance ele terá.

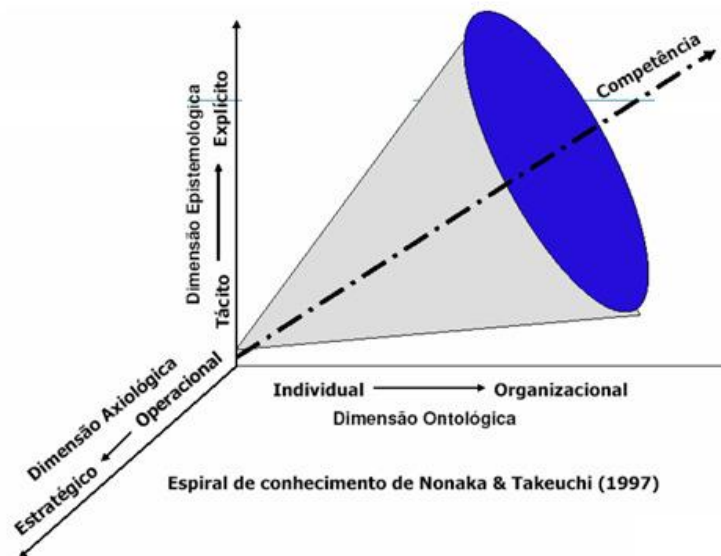


Figura 3. Espiral do conhecimento

Fonte: (Nonaka e Takeuchi, 1997)

2.1.2.3 MODELO DE PRAX

Prax (1997) afirma que a organização é o todo e este todo é superior à soma das partes, e afirma ainda que o conhecimento é o motor da dinâmica operacional da organização. A Figura 4 ilustra o modelo proposto por Prax no qual estipula três dimensões para a aplicação do paradigma do conhecimento melhor definido por Youssef e colaboradores (2006), como se segue:

- *Dimensão do homem e seu conhecimento:* nesta dimensão, define-se a comunicação em seu conceito complexo por meio dos grandes estágios da história da comunicação. Esses estágios, através de suas características, exercem uma grande influência sobre os modelos mentais e cognitivos dos agentes de comunicação. Na dimensão do homem, pode-se explicitar três variáveis: o conhecimento, a linguagem e a biografia.
- *Dimensão da empresa e do conhecimento organizacional:* composta por três variáveis (organização, estratégias e competências), esta dimensão está relacionada com os conceitos de conhecimento, sua comunicação e com a problemática do conhecimento coletivo, partindo da premissa de que as organizações se sentem ameaçadas pela nova realidade. Portanto, para que elas permaneçam competitivas, devem repensar a organização, revisando suas estratégias e reciclando as competências humanas.
- *Dimensão de novas tecnologias e a engenharia do conhecimento:* essa dimensão consiste em uma gama de recursos tecnológicos que contribuem com a engenharia do

conhecimento como, por exemplo: *grupoware workflow*, gestão e edição eletrônica de documentos.

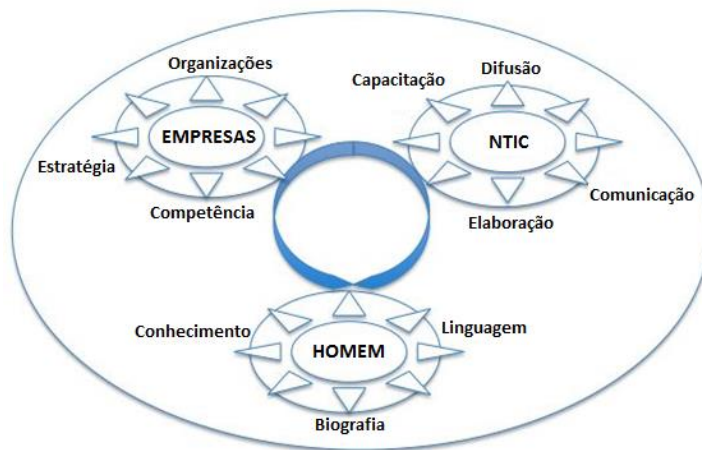


Figura 4. Modelo de GC de Prax

Fonte: Prax (1997)

2.1.2.4 MODELO DE CAVACANTI

O modelo de GC apresentado por Cavalcanti, Gomes e Neto (2001), é embasado em quatro pilares:

- *Capital Ambiental*: refere-se ao ambiente e todo o contexto organizacional.
- *Capital Intelectual*: refere-se ao ativo intangível em termos de conhecimento, ou seja, as pessoas que detêm o conhecimento que move a organização;
- *Capital de Relacionamento*: refere-se às alianças e às parcerias estabelecidas pela organização, assim como a interação entre os pilares envolvidos.
- *Capital Estrutural*: refere-se à estrutura organizacional e ao conjunto de sistemas que integram a organização.

A Figura 5 ilustra o modelo de gestão e principalmente a inter-relação existente entre os pilares do conhecimento, evidenciando o dinamismo e a mútua participação na criação do conhecimento.

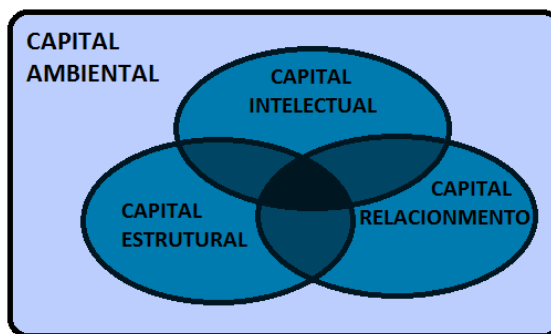


Figura 5. Modelo de GC de Cavalcanti, Gomes e Neto.

Fonte: Cavalcanti, Gomes e Neto (2001).

2.1.2.5 MODELO DE ANGELONI

O modelo de Angeloni (2002) traz um conjunto de regras por meio das quais se interpreta as realidades interna e externa do ambiente organizacional, constituindo-se de três dimensões interagentes e interdependentes, nomeadamente: infraestrutura organizacional, pessoas e tecnologia (Figura 6).

A dimensão infraestrutura organizacional contém aspectos e identidades da organização, tais como: cultura organizacional, estilo gerencial e estrutura organizacional, entre outros.

A dimensão pessoas refere-se ao capital intelectual, pois somente através do seu conhecimento científico e empírico podem-se criar novos conhecimentos oriundos de novas experiências, percepções e modelos mentais.



Figura 6. Modelo de Angeloni

Fonte: Adaptado de Angeloni (2002)

Finalmente, a dimensão tecnológica envolve o incremento de tecnologias e ambiente de GC no intuito de captar e distribuir conhecimentos na organização.

2.1.2.6 MODELO DE STOLLENWERK

Stollenwerk (2001) caracterizou um modelo baseado na análise e comparação de diversos modelos existentes na literatura, identificando assim sete processos de GC, a saber:

- Identificação de competências essenciais;
- Captura do conhecimento, habilidades e experiências para criar e manter as competências essenciais e áreas do conhecimento selecionadas e mapeadas;
- Seleção, validação e filtragem do conhecimento;
- Organização e armazenamento do conhecimento;
- Compartilhamento e disseminação do conhecimento;
- Aplicação e utilização do conhecimento; e
- Criação, como um processo que congrega várias outras dimensões, como aprendizagem, externalização do conhecimento e pesquisa.

A Figura 7 ilustra o esquema do modelo de GC de Stollenwerk, baseado segundo Miranda (2004), na adaptação da Ferramenta de avaliação de GC (*Knowledge Management Assessment Tool* - KMAT), tem como finalidade criar uma infraestrutura para suporte à: captura, análise, síntese, aplicação, distribuição e atribuição de valor ao conhecimento organizacional.



Figura 7. Modelo de GC de Stollenwerk
Fonte: STOLLENWERK (2001).

2.1.2.7 MODELO DE CHUN WEI CHOO (2006)

Choo (1998) descreveu um modelo de GC que enfatiza a criação do significado baseado em (WEICK, 2001), a criação do conhecimento baseado em (NONAKA e TAKEUCHI, 1995) e tomada de decisão baseado em (SIMON, 1957).

No modelo de GC de Choo a ação organizacional resulta da concentração e absorção de informação a partir do ambiente externo para cada ciclo sucessivo, como ilustrado na Figura 8. A cada fase, a criação de significado, a criação de conhecimento e a tomada de decisão têm um estímulo externo.

Na fase de tomada de decisão, analisam-se as informações que estão em fluxo do ambiente externo, identificam-se as prioridades para posterior filtragem de informações. Posteriormente, os indivíduos interpretam fragmentos de informação, combinadas com suas experiências anteriores.

O modelo proposto por Choo (2006) busca criar significado nas informações oriundas das atividades diárias, construir conhecimento e para, posteriormente, tomar decisões. Tais estruturas de processo estão intimamente ligadas e propiciam a GC.

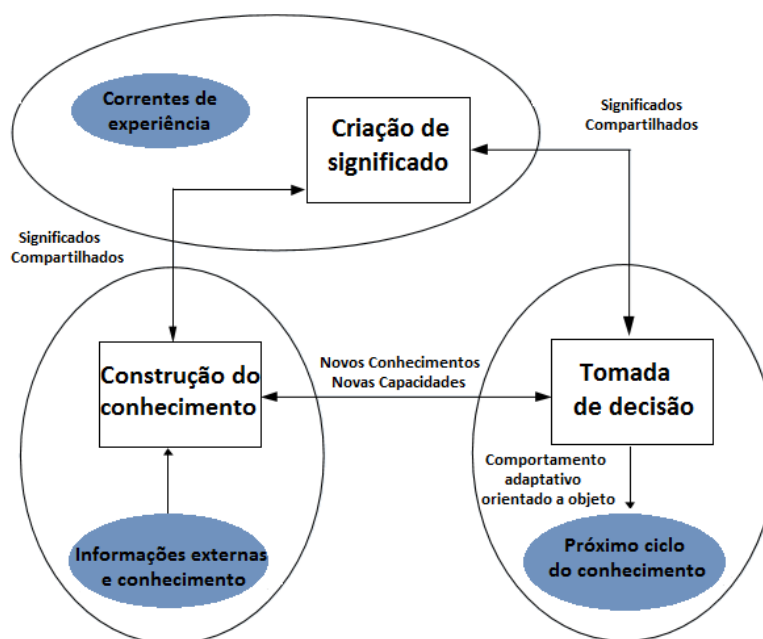


Figura 8. O ciclo do conhecimento. Fonte: Choo (2006).

Identifica-se na Figura 8 o ciclo do conhecimento do modelo Choo, conclui que a organização do conhecimento é capaz de integrar eficientemente os processos de criação de significado, construção do conhecimento e tomada de decisões. Segundo Choo (2006) é

necessário conhecimento para uma tomada de decisão, visto que quem detém o conhecimento possui vantagem competitiva, permitindo-lhe administrar os recursos, empenhar-se na aprendizagem constante, agir com inteligência e criatividade.

2.1.2.8 MODELO DE WIIG PARA DESENVOLVIMENTO E USO DO CONHECIMENTO

Wiig (1993) propõe uma hierarquia de conhecimento que consiste em formas (público, compartilhados e pessoais) e tipos de conhecimento (factual, conceitual, de expectativa e metodológico). Sua hierarquia de formas de conhecimento é mostrada na Figura 9. A principal vantagem do modelo Wiig é que, apesar de ter sido formulada em 1993, a abordagem organizada para categorizar o tipo de conhecimento a ser gerido continua sendo um conceituado modelo teórico de gestão de conhecimento. O modelo Wiig de gestão de conhecimento é talvez o mais pragmático dos modelos existentes na atualidade e pode ser facilmente integrado em qualquer das outras abordagens. Este modelo permite que os profissionais envolvidos no contexto organizacional adotem uma abordagem mais detalhada ou, mesmo, uma abordagem mais refinada de GC com base no tipo de conhecimento, mas vai além da dicotomia básica do conhecimento tácito/explicito.



Figura 9. Modelo de Wiig (WIIG, 1993)

O modelo de gestão de conhecimento de Wiig define diferentes níveis de internalização do conhecimento, podendo assim ser visto como um refinamento do modelo de Nonaka e Takeuchi.

A seguir a descrição da hierarquia do modelo de Wiig (Figura 9):

- Conhecimento público é explícito, geralmente disponível em domínio público rotineiramente compartilhado.
- Experiências compartilhadas são ativos de conhecimentos proprietários que são exclusivamente detidos por profissionais no âmbito organizacional e partilhadas no seu trabalho ou embutidos em tecnologia.
- Conhecimento pessoal é a forma menos acessível, no entanto a mais completa, do conhecimento.

Além destas três principais formas de conhecimento, Wiig (1993) define quatro tipos de conhecimento, nomeadamente factual, conceitual, de expectativa e metodológico, descritos como se segue:

- Conhecimento factual com dados e cadeias causais, medidas e conteúdos.
- Conhecimento conceitual envolve sistemas, conceitos e perspectivas.
- Conhecimento de expectativa diz respeito a julgamentos, hipóteses e expectativas dos conhecedores.
- Conhecimento metodológico lida com o raciocínio, com as estratégias, com os métodos de tomada de decisão, e outras técnicas.

Em conjunto, as três formas do conhecimento e os quatro tipos de conhecimento se combinam para produzir uma matriz de gestão de conhecimento que constitui a base do modelo Wiig de gestão de conhecimento.

2.1.2.9 MODELO DE GC DE BOISOT I-SPACE

O modelo GC de Boisot apresenta uma abordagem tácita do conhecimento, observando que em muitas situações, após a codificação, ocorre uma perda de contexto o que pode resultar na perda de conteúdo valioso. Este conteúdo necessita de um contexto compartilhado para a sua interpretação, isto implica interação e socialização como se pode observar no modelo de Nonaka e Takeuchi (1995).

O modelo I-Space pode ser visualizado como um cubo tridimensional com as seguintes dimensões: (1) codificado – não codificado; (2) concreto – abstrato; e (3) difuso não difuso.

O modelo Boisot incorpora uma base teórica da aprendizagem social e serve para unir conteúdo, informação e GC. Em um sentido aproximado, a dimensão de codificação está ligada à categorização e à classificação; a dimensão de abstração está ligada à criação de conhecimento através da análise e compreensão; e a terceira dimensão de difusão está ligada ao acesso e transferência de conhecimento. Existe um forte potencial para fazer uso do modelo gestão de conhecimento Boisot I-Space para mapear e gerenciar ativos de conhecimento de uma organização, assim como para a aprendizagem social e o ciclo de algo que os outros modelos de GC não abordam diretamente. No entanto, o modelo Boisot parece ser pouco conhecido e menos acessível e, como resultado, não tem tido aplicação generalizada. Mais atividade no campo de testes deste modelo iria fornecer *feedback* sobre a sua aplicabilidade bem como mais orientações sobre a melhor maneira de implementar a abordagem I-Space.

2.1.2.10 MODELO DE GC DE SISTEMA COMPLEXO ADAPTATIVO INTELIGENTE

O sistema complexo adaptativo inteligente (Intelligent Complex Adaptive Systems - ICAS) consistem em agentes⁴ que interagem uns com os outros localmente, que apresentam comportamento combinado propiciando fenômenos complexos adaptativos. Não há nenhuma gerencia sobre agentes, atuando de forma independente. Desta forma um padrão geral de comportamento complexo surge como o resultado de todas as suas interações. Por este motivo os sistemas adaptativos complexos são ditos "auto organizáveis" (BEER, 1981)(BENNET e BENNET, 2004).

Este modelo enfatiza o conhecimento individual do trabalhador, assim como suas competência e capacidade de aprendizagem. Estes ativos de conhecimento são aproveitados através de múltiplas redes para disponibilizar o conhecimento e a experiência. Desta forma, para que uma organização possa ter acesso aos ativos do conhecimento são necessárias oito características emergentes, nomeadamente: (1) inteligência organizacional; (2) propósito compartilhado; (3) selectividade; (4) a complexidade aceitável; (5) fronteiras permeáveis; (6) centralização do conhecimento; (7) fluxo; e (8) multidimensionalidade

As características emergentes ICAS são apresentadas na Figura 10. Estas características servem para dotar a organização com a capacidade interna para lidar com os futuros ambientes inesperados ainda a serem encontrados. A inteligência organizacional

⁴ Agente: é algo capaz de perceber seu ambiente por meio de sensores e de agir sobre esse ambiente por meio de atuadores (RUSSEL e NORVIG, 2004)

refere-se à capacidade da organização de inovar, adquirir conhecimento e aplicar esse conhecimento a situações relevantes.

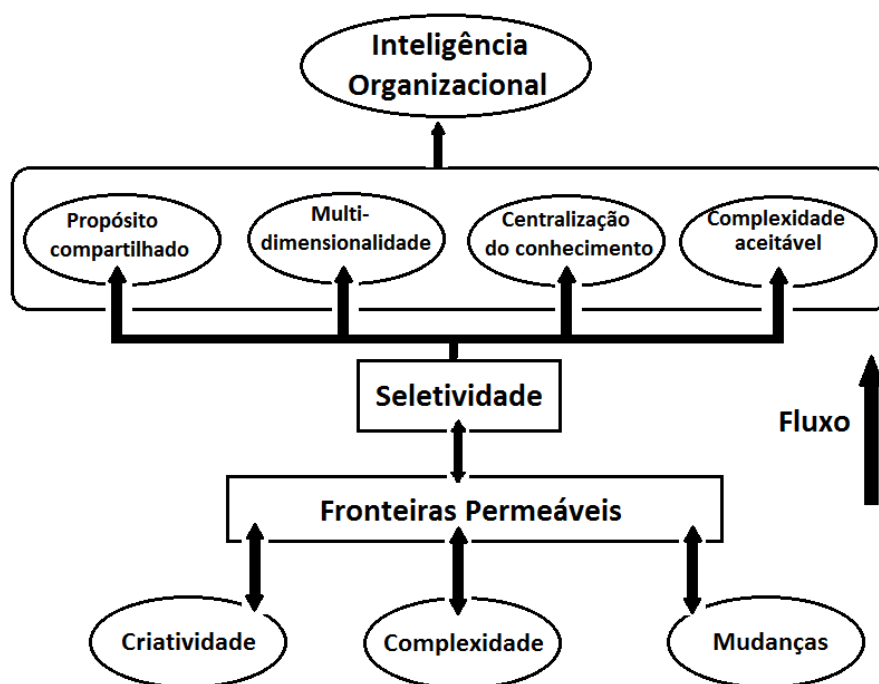


Figura 10. Modelo ICAS (Bennet e Bennet, 2004)

O processo de seletividade consiste em filtrar informações recebidas do mundo exterior. No entanto, para que ocorra uma “boa filtragem” é necessário um amplo conhecimento da organização, o conhecimento específico do cliente, e uma forte compreensão dos objetivos estratégicos da organização. A centralidade do conhecimento refere-se à agregação de informações relevantes de auto-organização, colaboração e alinhamento estratégico. Fluxo permite o foco no conhecimento e facilita as conexões e a continuidade necessárias para manter a unidade e dar coerência à inteligência organizacional. Fronteiras permeáveis são essenciais para que as idéias possam ser trocadas e construídas. Finalmente, a multidimensionalidade representa flexibilidade organizacional que garante que os trabalhadores do conhecimento tenham a competência, perspectiva e capacidade cognitiva para tratar de questões e resolver problemas.

2.1.2.11 PERCEPÇÃO DOS MODELOS DE GESTÃO DO CONHECIMENTO

Os modelos de gestão do conhecimento apresentados neste capítulo condensam um arcabouço teórico dos processos e ciclos de vida da gestão do conhecimento, dispondo desta forma uma base sólida para compreensão mais profunda do que é gestão do conhecimento.

A observação e a análise de diferentes abordagens de modelos de gestão do conhecimento permite mapear diferentes aplicações das práticas de GC, que por sua vez enquadra-se em um contexto ou problemática específica no âmbito da formalização do conhecimento que trata este trabalho.

Optou-se neste trabalho aplicar o modelo de Nonaka e Takeuchi como base para formalização do conhecimento, considerando as restrições existentes sobre os tipos de conhecimento: explícito e tácito. Entende-se que a distinção entre esses dois tipos de conhecimento se dá pela forma como o conhecimento é representado. No entanto, não há como trata-los de forma separada, pois, de alguma maneira, categoricamente o conhecimento explícito vem acompanhado do conhecimento tácito (NISSEN; KAMEL; SENGUPTA, 2000). Isto ocorre em detrimento de que "as pessoas e a cultura do domínio que estão inseridas são os fatores de condução que, determinam o sucesso ou o fracasso das práticas de gestão do conhecimento" (RUBENSTEIN-MONTANO et al., 2001).

Neste sentido, o entendimento dos modelos de gestão do conhecimento permite conceber elementos fundamentais neste trabalho, a citar: Ontologia, vetor semântico, taxonomia de conceitos.

2.1.4 Referencial Semântico: A Ontologia

O Referencial Semântico situa-se como parte integrante da GC e tem como finalidade propiciar recursos que viabilizem o acesso ao conhecimento. Dentre as tecnologias semânticas que abrange, destaca-se: taxonomia de conceitos, vetores semânticos e ontologias. Dentre estes o mais referenciado na literatura é a ontologia.

Segundo Guarino (1998):

“no sentido filosófico, podemos nos referir a uma ontologia como um sistema particular de categorias, responsável por uma determinada visão do mundo. Como tal, este sistema não depende de um idioma em particular: a ontologia de Aristóteles é sempre a mesma, independentemente do idioma usado para descrevê-la. Por outro lado, no seu uso mais predominante na Inteligência Artificial, uma ontologia refere-se a um artefato de engenharia, constituído por um vocabulário específico usado para descrever uma dada realidade mais um conjunto de pressupostos explícitos a respeito do significado pretendido para as palavras do vocabulário”.

Basicamente, o papel da ontologia é facilitar a construção de um modelo de domínio por meio da representação de um vocabulário de conceitos e relações (STUDER, DECKER et

al., 2000). Guarino (1998) define ontologia como sendo *uma teoria lógica que pretende representar ou explicar um determinado significado por meio de um vocabulário formal*. Mais precisamente, conforme discutido por Chandrasekaran e colaboradores (1999), não é esse vocabulário que por si só define uma ontologia, mas sim as conceitualizações que os termos descritos nesse vocabulário pretendem capturar. Além disso, a formalização de uma ontologia é dependente de uma linguagem, ao passo que uma conceitualização independe de uma linguagem (GUARINO, 1998).

Segundo Guarino (1998, p.2):

Uma ontologia se refere a um artefato de engenharia (de software), que é constituído por um vocabulário específico utilizado para descrever certa realidade, mais um conjunto de suposições explícitas a respeito do significado pretendido para as palavras do vocabulário. Esse conjunto de suposições tem, em geral, a forma da teoria da lógica de primeira ordem, onde palavras do vocabulário aparecem com nomes de predicados unários ou binários, respectivamente chamados conceitos e relações. No caso mais simples, uma ontologia descreve uma hierarquia de conceitos relacionados por relações de classificação; em casos mais sofisticados, axiomas são adicionados à estrutura de forma a expressar outras relações entre conceitos, e para restringir a interpretação pretendida para tais conceitos.

Segundo Studer e colaboradores (1998) pode-se afirmar que uma ontologia é uma especificação de conhecimento consensual sobre um modelo abstrato de domínio, definida explicitamente em termos de conceitos, suas propriedades e relações, por meio de axiomas, possibilitando, assim, que seja automaticamente interpretado por programas computacionais.

2.1.4.1 TIPOS DE ONTOLOGIA

Observa-se na literatura várias formas de classificações de ontologias expressas por diversos autores, sendo que uma das mais utilizadas é a de Guarino (1998), que utiliza características-chaves das ontologias, sugerindo o desenvolvimento de diferentes tipos de ontologia de acordo com o nível de generalidade necessária. No entanto Maedche (2002) complementa a divisão atribuindo quatro tipos de ontologias, são as seguintes:

- *Alto nível*: descrevem conceitos gerais, que não são particulares a um domínio ou problema específico. Feitas para serem utilizadas por uma gama diversificada de usuários.

- *Domínio*: descrevem o vocabulário relativo a um domínio, de forma genérica. Podem especializar termos das ontologias de alto nível.
- *Tarefa*: descrevem o vocabulário relativo a uma tarefa ou atividade, de forma genérica. Podem especializar termos das ontologias de alto nível.
- *Aplicação*: descrevem conceitos correlatos tanto a um domínio específico quanto a uma tarefa em particular. Estes conceitos são, em geral, papéis desempenhados por entidades presentes no domínio que executam determinada atividade.

2.1.4.2 COMPOSIÇÃO DA ONTOLOGIA

Uma ontologia pode ser estruturada de várias formas, mas necessariamente inclui um vocabulário de termos e alguma especificação de seus significados (USCHOLD e GRUNINGER, 1996). O nível de formalidade de uma ontologia pode variar, mas para os propósitos desta tese consideram-se ontologias formais, cujos termos sejam definidos com semântica formal. Os componentes básicos de uma ontologia são classes, relações, axiomas e instâncias. Gruber (1993) descreve estes componentes da seguinte forma:

- *Classes*: também chamadas comumente de conceitos, podem ser do tipo abstrato ou concreto, simples ou composto, reais ou fictícios. Em suma, um conceito pode ser “qualquer coisa” a respeito de “algo” que está sendo explicando, e por esse motivo pode ser a descrição de uma tarefa, função, ação, estratégia ou um processo de raciocínio.
- *Relações e funções*: relações são um tipo de interação entre as classes de um domínio e seus atributos. Já as funções são um tipo especial de relação (e.g., Exponencial (x)).
- *Axiomas*: utilizados para modelar sentenças que são sempre verdadeiras. Os axiomas podem ser utilizados para vários fins, tais como: impor restrições, verificar a correção e realizar dedução de novas informações. Em outras palavras, os axiomas são usados para restringir a interpretação e o uso dos conceitos envolvidos na ontologia.
- *Instâncias* (ou *Indivíduos*): representam elementos do domínio associados a um conceito específico. As instâncias possuem atributos que são propriedades relevantes que descrevem a individualidade de um conceito.

2.1.4.3 DESENVOLVIMENTO DA ONTOLOGIA

Para Noy e McGuinness (2001), o desenvolvimento de uma ontologia inclui os seguintes passos:

- Definir as classes da ontologia;

- Arranjar as classes em uma hierarquia (subclasses e superclasses);
- Definir as propriedades das classes e descrever a gama de valores válidos para elas; e
- Preencher os valores de propriedades para as instâncias.

Entretanto, elaborar ontologias envolve mais do que apenas definir os conceitos da ontologia de maneira formal. Deve-se determinar o escopo do domínio, analisá-lo para capturar a sua conceituação, considerar o reuso de ontologias já existentes, entre outras atividades como formas de representação, armazenamento e avaliação. Há diversas propostas de processos de desenvolvimento de ontologias, como em Uschold e Gruninger (1996), Noy e McGuinness (2001). O último se compõe das melhores práticas de outras metodologias e está dividido em cinco grandes atividades, nomeadamente:

- *Especificação*: avaliar os custos do desenvolvimento da ontologia.
- *Conceituação*: descrever um modelo conceitual do domínio de discurso.
- *Formalização*: transformar o modelo conceitual em um modelo formal, passível de ser implementado.
- *Implementação*: implementar a ontologia formalizada em uma linguagem de representação adequada
- *Avaliação*: a ontologia é validada quanto ao entendimento aceito sobre o domínio em fontes de conhecimento. Verifica-se a coerência do conhecimento representado na ontologia e certifica-se de sua utilidade.

2.1.5 Vetor Semântico

Segundo Jurafsky e Martin (2015), vetores semânticos são geralmente baseados em uma matriz de co-ocorrência, uma forma de representar: (i) a frequência com que ocorrem palavras; (ii) os pesos que estas palavras representam em determinado contexto; e (iii) outras características que a linguística permite. Vetor semântico é uma matriz de termos de uma fonte linguística (e.g. documentos, ontologia) com representação matemática o que permite raciocinar semanticamente, além de apoiar aplicações estatísticas e probabilísticas.

A criação de um vetor semântico é apoiada pelo Modelo de Espaço Vetorial (do inglês: Vector Space Model - VSM) que oferece abordagens para implementar a extração de conhecimento a partir de fontes (de conhecimento), assim como representar o conhecimento de forma organizada para análise de relevância entre os conceitos de um base de conhecimento.

O VSM é um modelo matemático. Um modelo matemático capaz de representar os significados dos itens lexicais como vetores em um "espaço semântico" (SALTON et. al., 1975). O benefício dos VSMs é que eles podem facilmente ser manipulados utilizando álgebra linear, permitindo certo grau de semelhança entre os vetores.

Modelos vetoriais semânticos são modelos em que os conceitos são representados por vetores em um espaço dimensional. Similaridade entre conceitos pode ser calculada usando a analogia de similaridade ou distância entre os pontos neste espaço vetor. Modelos vetoriais semânticos são agora uma parte reconhecida da linguística computacional e são por vezes descritos como modelos *wordspace* (WIDDOWS, 2004)(SAHLGREN, 2006).

Os VSMs incluem uma família de modelos relacionados para representar conceitos com vetores em um espaço vetorial de dimensões elevadas, como a análise semântica latente (LANDAUER e DUMAIS, 1997), grandes espaços analógicos para a linguagem (LUND e BURGESS, 1996), e WORDSPACE (SCHÜTZE, 1998)(WIDDOWS, 2004)(SAHLGREN, 2006).

Ao longo de vários anos, os pontos fortes destes modelos vetoriais semânticos foram examinados e avaliados no desempenho de várias tarefas de importância para o processamento da linguagem natural. Tais aplicações de VSMs incluem:

- Recuperação da informação (DEERWESTER et al., 1990): como aplicações de redução de dimensão para uma matriz de termos de um documento, visando criar um motor de busca semanticamente ciente (por exemplo, um motor de busca que pode localizar documentos com base em sinônimos e termos relacionados, bem como palavras-chave correspondentes).
- Interação semântica - ontologia (HEARST e SCHÜTZE, 1993)(WIDDOWS, 2003): o princípio fundamental aqui é que o conhecimento de algumas descendências (hierarquia) de palavras e seus relacionamentos pode ajudar a inferir relações análogas para outras palavras similares que estão nas proximidades no espaço vetorial semântico.
- Segmentação do documento (BRANTS et al., 2002): dado que podemos calcular vetores de contexto para as regiões do texto, é possível detectar as discordâncias de contexto, quando ocorrem mudanças de áreas de texto ou mesmo em mudanças de um documento para outro.

2.1.5.1 ESPAÇO VETORIAL EUCLIDIANO

Um espaço vetorial euclidiano é definido pelos cinco postulados de Euclides, que são a base para a Geometria Clássica:

- Uma linha reta, ou vetor, pode ser desenhado de qualquer ponto para qualquer outro ponto (2 pontos determinam um vetor único).
- Um vetor finito pode ser produzido com qualquer comprimento em uma linha reta.
- Um círculo pode ser descrito com qualquer centro a qualquer distância desse centro.
- Todos os ângulos direitos são iguais
- Caso um vetor encontre com dois outros vetores, de modo que os dois ângulos interiores sejam menores que um ângulo reto, os outros vetores se encontrarão, se somente se produzidos no mesmo lado dos inferiores ao ângulos retos. Considerando \mathbb{R}^n como o domínio de um espaço vetorial euclidiano com n dimensões e definindo vetores $x, y \in \mathbb{R}^n$ como:

$$x = (x_1, x_2, \dots, x_n) \quad y = (y_1, y_2, \dots, y_n)$$

Então, a distância euclidiana entre x , mostrada na Figura 11, é definida como o produto interno entre x e y , e dado por (DEZA, 2009):

Na Equação 1:

$$\cos \theta = \frac{\text{"lado adjacente"}}{\text{"hipotenusa"}} = \frac{x \cdot y}{\|x\| \|y\|}$$

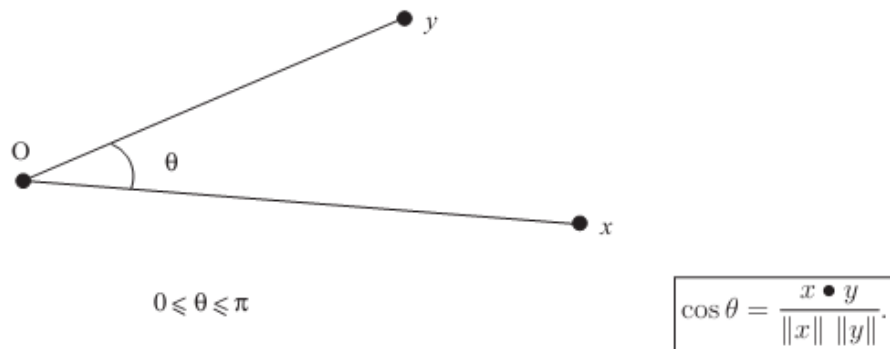


Figura 11. Distância euclidiana entre dois vetores

Definição 1: Para qualquer $y \in \mathbb{R}^n$, o produto interno de x e y , também conhecido como produto ponto, é o número

$$x \cdot y = \sum_{i=1}^n x_i y_i$$

Definição 2: A norma de um vetor $x \in \mathbb{R}^n$ é o número

$$\|x\| = \sqrt{\sum_{i=1}^n x_i^2}$$

Conclui-se, a partir da equação, que a distância euclidiana determina que os comprimentos dos vetores devam corresponder. Isto implica um problema com a comparação entre vetores que não têm o mesmo comprimento. No entanto, existem alguns casos em que a distância Euclidiana pode ser usada para calcular a distância entre vetores com comprimentos diferentes, com o uso de Multiplicação Esparsa-Matriz, como mostrado na subseção seguinte.

2.1.5.2 MULTIPLICAÇÃO DE MATRIX ESPARSA

Como se pode compreender obviamente, vetores semânticos não têm necessariamente o mesmo tamanho. Isto significa que a distância Euclidiana não pode ser aplicada diretamente em vetores semânticos. Para calcular a função cosseno entre dois vetores com diferentes tamanhos, deve-se usar uma aproximação de multiplicação de matriz-separata.

Uma matriz esparsa é uma matriz com apenas uma pequena porcentagem de valores não nulos. Ao multiplicar duas matrizes esparsas, estas podem ter tamanhos diferentes, pois valores nulos podem ser adicionados ao menor vetor, tornando-o do mesmo tamanho que o maior vetor. Especificamente, considerando dois vetores $x \in \mathbb{R}^n$ e $y \in \mathbb{R}^m$, com $m > n$, como:

$$x = (x_1, x_2, \dots, x_n) \quad y = (y_1, y_2, \dots, y_m)$$

Para realizar a multiplicação de x por y , o tamanho do vetor x tem de ser aumentado, adicionando $m-n$ valores nulos a x . O sistema SEKS utiliza esta abordagem de multiplicação de matriz esparsa devido ao fato de que a medida de distância interna do produto para os vetores x e y apresentados acima pode ser decomposta em três valores: um dependendo dos valores não nulos de x , $f_2(x_i)$, outro dependendo dos valores não nulos de y , $f_3(y_i)$, e o terceiro dependendo das coordenadas não nulas compartilhadas por x e y , $f_1(x_i, y_i)$. Formalmente:

$$\cos \theta = f_0 \left(\sum_{k=1}^m f_1(x_i, y_i), f_2(x_i), f_3(y_i) \right)$$

Na equação 2, $f_0, f_1(x_i, y_i), f_2(x_i)$ e $f_3(y_i)$ são dados, respectivamente, por:

$$f_0(a, b, c) = \frac{a}{bc}$$

$$f_1(a, b) = ab$$

$$f_2(a) = f_3(a) = \sqrt{\sum_{k=1}^m a_k^2}$$

As quatro expressões definidas acima podem ser combinadas para realizar a Equação 1. Neste caso, embora o método do produto interno inicialmente exija que ambos os vetores tenham o mesmo tamanho, ao aplicar a abordagem de multiplicação de matriz esparsa apresentada acima, os tamanhos dos vetores não necessariamente podem coincidir.

Se um vetor for menor que o outro, então, significa, que o vetor menor tem valores nulos para todos os conceitos que estão faltando para alcançar o tamanho do vetor maior. Por outro lado, o cálculo de $f_1(x_i, y_i)$ só é necessário quando ambos os vetores têm pelo menos uma coordenada não nula compartilhada. Se os vetores não possuem qualquer conceito compartilhado, isto é, uma coordenada diferente de zero, o valor para a função acima é nulo, e os vetores não apresentam qualquer semelhança. Isso também significa que f_2 e f_3 não precisam ser calculados, reduzindo significativamente a computação necessária (TURNERY e PANTEL, 2010).

2.2 Os Sistemas Especialistas (SEs)

A Inteligência Artificial é uma área da ciência da computação que busca, através de técnicas inspiradas na natureza, o desenvolvimento de sistemas inteligentes que imitam aspectos do comportamento humano, tais como aprendizado, percepção, raciocínio, evolução e adaptação (PACHECO, 2002).

O termo surgiu na década de 60 e, no início das pesquisas os cientistas Newell, Simon, e J. C. Shaw introduziram o processamento simbólico. Ao invés de construir sistemas baseados em números, eles tentaram construir sistemas que manipulassem símbolos. A abordagem era poderosa e foi fundamental para muitos trabalhos posteriores, porém era muito ambiciosa, visavam construir “resolvedores de problemas” genéricos com interface em linguagem natural.

Posteriormente, as pesquisas em inteligência artificial começaram a obter melhores resultados quando passaram a focar em problemas mais restritos, envolvendo áreas específicas do conhecimento. Os sistemas capazes de abstrair o conhecimento de um

especialista na realização de uma atividade foram denominados Sistemas Especialistas (SEs) (HAYES-ROTH, 1983).

SEs são tradicionalmente definidos como programas computacionais que modelam o conhecimento e emulam o processo de raciocínio de um especialista humano na resolução de um problema específico de um determinado domínio (TAYLOR e LUBKEMAN, 1989)(SUSTAETA et al, 1989)(LAVALLE e RODRIGUEZ, 1989).

Os SEs, quando aplicado em domínios restritos, podem apresentar desempenhos próximos ao desempenho de um indivíduo especialista. No entanto observa-se no final da década de 80, mudanças no conceito de SEs que passam a ser caracterizados como ferramenta de suporte à tomada de decisão em processos complexos e exaustivos, propiciando desta forma alternativas e perspectivas para usuário, responsável pelo julgamento final (CHENG, 1988)(HWA,1987)(LIMA,1988).

Neste sentido, alguns autores preferem denominar sistemas projetados com tal objetivo de "sistemas baseados em conhecimento", ou simplesmente de "sistemas de conhecimento" (CHENG, 1988)(HWA, 1987).

Posteriormente, surgiram outras formas de representação de sistemas baseados em conhecimento, como por exemplo: Sistemas de Processamento de Linguagens Naturais e os Sistemas de Redes Neurais Artificiais.

Desta forma, os SEs são Sistemas Baseados em Conhecimento (SBC), mas a recíproca não é necessariamente verdadeira, pois nem todos os Sistemas Baseados em Conhecimento podem ser classificados como SEs (WATERMAN, 1986).

Waterman (1986) representa a relação entre SBCs e os SEs:

- Sistemas Baseados em Conhecimento: são sistemas onde o domínio do conhecimento é explícito e separado do restante do sistema;
- SEs: são sistemas que aplicam o conhecimento especializado na resolução de problemas do mundo real.

Segundo Souto (2005) os SEs diferem dos sistemas convencionais por solucionarem problemas, procurando traduzir a estrutura de pensamentos dos especialistas humanos, usando, para tanto, estruturas de conhecimento e heurísticas, enquanto que os sistemas convencionais procuram soluções para os problemas, através de modelos algorítmicos.

2.2.1 Classificação dos Sistemas Especialistas

De um modo geral, os SEs são classificados quanto às características do seu funcionamento. Tais características são provenientes da metodologia de resolução e do tipo de problema que o sistema será empregado. Segundo Hayes-Roth e colaboradores (1983) destacam-se dentre as aplicações de SEs:

- *Interpretação*: consiste na análise de dados e atribuição de significados simbólicos, na prática da interpretação de objetos a partir de conjuntos de observações (compreensão de fala, análise de imagens, interpretação geológica - e.g. análise de imagens de carro para reconhecimento de placas).
- *Classificação e Diagnóstico*: a classificação é o processo de extração de informação e de características que sirvam de artifício para o reconhecimento padrão ou para a separação de indivíduos afins. A tarefa de diagnóstico é análoga à classificação, visa identificar respostas baseadas em sintomas e características de um domínio ou objeto de aplicação (e.g. diagnósticos médicos, mecânicos).
- *Projeto*: consiste no desenvolvimento de especificações e configurações de objetos que satisfazem determinados requisitos ou restrições (e.g. projeto de circuitos digitais, projeto de edifícios).
- *Monitoramento*: consiste em comparar observações de comportamento de sistemas com características desejáveis e, em caso de anormalidade, determinar um conjunto de ações necessárias para alcançar objetivos (e.g. monitoração de rede de distribuição de energia elétrica, controle de tráfego aéreo).
- *Controle*: trata de ambientes de automação industrial no qual consiste em determinar um conjunto de ações, baseado em regras e no comportamento do sistema (e.g. robôs, gerência de produção).
- *Avaliativo*: consiste na prática de analisar, averiguar, medir, comparar e inferir sobre determinada ação de forma automática, baseada na experiência avaliativa de um grupo de especialistas com intuito de reutilizá-la de forma padronizada (e.g. avaliação de software de automação industrial, avaliação de qualidade de energia).

2.2.2 Conhecimento Especialista

Segundo Weiss, Kulikowski (1988) e Prado (2001), um especialista humano tem tipos diferentes de informação para fornecer ao criador de um modelo de raciocínio especialista, incluindo: experiência pessoal na solução de problemas, perícia pessoal ou métodos para a

solução de problemas e conhecimento pessoal sobre as razões da seleção dos métodos utilizados.

Relacionar a experiência pessoal com a solução de problemas, estabelecendo regras resumidas de habilidade, raramente constitui-se uma tarefa fácil para os especialistas humanos. Com frequência, os especialistas são pressionados não só para descreverem sua perícia de maneira sistemática mas, também, sob forma racionalmente estruturada.

O conhecimento especializado, além de representar a estrutura do domínio específico, representa um contexto histórico. A produção de conhecimento é uma atividade dinâmica e constante, o que pode ocasionar modificações no conhecimento especializado, com o passar do tempo. Neste sentido é pertinente que um sistema especialista apresente flexibilidade para atualização do conhecimento, um suporte para integração do novo conhecimento ao conhecimento existente, apoiando, portanto, a transferência de conhecimento.

2.2.2.1 AQUISIÇÃO DE CONHECIMENTO (AC)

De acordo com Drumond e Girardi (2010), tradicionalmente a tarefa de construção das bases de conhecimento⁵ tem sido realizada manualmente por especialistas de domínio e por profissionais da área de computação (especializados na área de GC), o que aumenta a propensão a erros. Tal dificuldade em explicitar o conhecimento implícito nos textos e nas bases de dados é chamado de “aquisição de conhecimento” e superar esse problema é crucial para o sucesso de aplicações baseadas em conhecimento.

Dentre os processos que envolvem o desenvolvimento de um sistema especialista, o processo de aquisição do conhecimento destaca-se como prioritário, visto que caso ocorra excesso ou excasses de conhecimento o contexto semântico torna-se obsoleto ou volumoso (PRADO, 2001). Segundo Polanyi (1983), assume-se que as pessoas sabem muito mais do que conseguem falar ou transmitir. Desta forma, o processo de aquisição de conhecimento com pessoas especializadas em um domínio geralmente apresenta alto custo e elevado grau de complexidade.

A aquisição de conhecimento pode ser definida como o processo de compreender e organizar o conhecimento de várias fontes (MASTELA, 2004). Esse conhecimento deverá ser codificado e armazenado em uma base de conhecimento para posterior resgate por um SE.

⁵ Base de Conhecimento: parte integrante dos sistemas de gestão do conhecimento, uma base de conhecimento tem como estrutura a ontologia de um domínio específico, por vezes é usado para otimizar a coleta de informações, organização e recuperação de uma organização (BUNGE, 2003).

O agente responsável por todas as atividades de construção de um SE chama-se engenheiro de conhecimento, que tem como função abstrair o conhecimento dos especialistas, e posteriormente traduzido para regras. Depois que o sistema inicial estiver pronto, ele precisa ser iterativamente refinado até aproximar-se do nível de desempenho de um especialista (RICH e KNIGHT, 1993).

2.2.2.2 TÉCNICA DE ELICITAÇÃO DO CONHECIMENTO

A elicitação é o processo de extração do conhecimento das fontes disponíveis. A pesquisa no âmbito da estruturação de conhecimento tem direcionado esforços para sistematizar ou até mesmo automatizar o processo de aquisição de conhecimento. Desta forma, as técnicas podem ser classificadas em manuais, semi-automáticas e automáticas.

A maioria das técnicas manuais fundamenta-se na psicologia e na análise de sistemas (SOLANGE et al., 2003). Nessas técnicas o engenheiro de conhecimento é responsável por adquirir o conhecimento do especialista e outras fontes de conhecimento para posteriormente codificá-lo na base de conhecimento. A aquisição de conhecimento semi-automática consiste na utilização de ferramentas computacionais que auxiliem ao engenheiro de conhecimento a codificação da base de conhecimento. Já as técnicas automáticas de aquisição de conhecimento dizem respeito ao processo pelo qual o conhecimento é adquirido automaticamente por mineração de dados e o aprendizado de máquina (redes neurais, árvores de decisões, entre outros).

Sobre as técnicas para a aquisição do conhecimento, Garcia, Varejão e Ferraz (2005) apresentam cinco categorias:

- Manuais baseadas em entrevistas, em modelos ou em acompanhamento;
- Semiautomáticas baseadas em teorias cognitivas ou em modelos que já existem;
- Que utilizam aprendizado de máquina tentando induzir regras a partir de exemplos catalogados;
- Que utilizam mineração de dados, a partir da qual se busca extrair regras e comportamentos com base em análises de grandes massas de dados; e
- Que aplicam mineração de texto para extrair o conhecimento de uma grande quantidade de dados não estruturados.

Outra abordagem é observada em (CORDINGLEY, 1989)(SHADBOLT, O'HARA e CROW, 1999) que define como técnica de obtenção de conhecimento: análise de mídias, análise de comportamento, cenários e entrevistas.

Na análise de mídias, realiza-se a extração de conhecimento por análise de textos, bases de conhecimento, diagramas e outras fontes de conhecimentos explícitos. A imersão na literatura facilita posteriores aquisições de conhecimento com o próprio especialista (CORDINGLEY, 1989).

A análise de comportamento ou técnica de observação consiste em acompanhar o especialista em sua rotina de trabalho, observando o comportamento e a execução de suas tarefas. Essa técnica, por utilizar casos reais, evita que o especialista seja direcionado a responder questões irrelevantes, no entanto, nem sempre se consegue uma amostragem de casos realmente representativa e em certos casos a abstração torna-se complexa (SHADBOLT, O'HARA e CROW, 1999).

O método consiste em analisar o processo de raciocínio do especialista em casos (tarefas) reais ou hipotéticos submetidos pelo elicitante. O método utiliza-se da teoria do raciocínio baseado em caso no qual busca resolver novos problemas adaptando soluções utilizadas para resolver problemas anteriores. Os casos devem representar o domínio, desta forma devem ser problemas relevantes, problemas que cubram as exceções e problemas de variados graus de imprecisão (CORDINGLEY, 1989).

Segundo Shadbolt, O'hara e Crow (1999) a entrevista é uma atividade de interação entre o elicitante e o especialista, que baseia-se em uma estratégia de perguntas e respostas, compondo uma base de informações. Essas informações são posteriormente analisadas para se extrair o conhecimento desejado. Há três tipos de entrevistas:

- *Estruturadas*: são entrevistas formais, a partir de uma relação fixa de perguntas, que envolvem pré-planejamento cuidadoso das questões e da ordem destas, bem como a especificação de eventos que o entrevistador deve e não deve fazer.
- *Semiestruturadas*: são livres e informais mas apresentam certo grau de estruturação, já que se guiam por uma relação de pontos de interesses que o entrevistador vai explorando ao longo do seu curso. O entrevistador faz poucas perguntas diretas e deixa o entrevistado falar livremente à medida que se refere às pautas assimiladas. Quando este, por ventura, se afasta, o entrevistador intervém de maneira sutil, para preservar a espontaneidade da entrevista.
- *Não Estruturadas*: só se distingue da simples conversação porque tem como objetivo básico de aquisição de conhecimento. O objetivo dessas entrevistas não é a adquirir conhecimento sobre um tópico específico, mas sim obter uma visão geral do domínio

em questão. Assim as entrevistas são mais livres e propiciam um bom relacionamento entre o engenheiro de conhecimento e o especialista.

2.2.2.3 REPRESENTAÇÃO DO CONHECIMENTO EM SES

A representação do conhecimento é o método usado para modelar o conhecimento de especialistas em algum domínio de aplicação, compondo uma estrutura de representação que permite a interação com os mecanismos de inferência dos sistemas inteligentes. A representação do conhecimento consiste em um arranjo entre estrutura de dados e procedimentos interpretativos, o resultado desta integração gera uma conduta inteligente (CUNHA, 1995).

Para Nilsson (1980) representar conhecimento implica em encontrar as estruturas adequadas para expressar o tipo de conhecimento particular do domínio da aplicação.

Sowa (2000) qualifica a representação do conhecimento como uma aplicação lógica na tarefa de construir modelos computacionais sobre algum domínio específico. O campo da representação do conhecimento geralmente é chamado de “representação do conhecimento e raciocínio”, pois os formalismos da representação do conhecimento tornam-se inúteis se não houver a possibilidade de raciocínio e inferência sobre eles.

O engenheiro do conhecimento não realiza somente a aquisição do conhecimento; cabe a ele formalizar a estrutura do conhecimento com formas de representação (RUSSELL e NORVIG, 2004).

2.2.2.4 Técnicas de Representação do Conhecimento

Existem várias técnicas para representação do conhecimento, dentre estas destacam-se as ontologias que não somente representam o conhecimento em um arcabouço semântico, mas possibilitam a reutilização e transmissão deste, além de ser uma forma estruturada para o seu armazenamento com a utilização do conceito de classes, relações e atributos (GÓMEZPÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004).

Segundo Luger (2009), o uso da lógica, regras, redes semânticas e frames são consideradas as principais alternativas para a representação de conhecimento:

- *Lógica*: utiliza-se de linguagem matemática de cálculo de predicados de primeira ordem. O método lógico representa o conhecimento através de sentenças lógicas que representam uma linguagem de representação formal com regras de inferências baseadas em deduções consistentes e completas, utiliza-se de operadores lógicos como “v” (ou) e “→” (então) (LUGER, 2009)

- *Regras*: o uso de regras para a representação do conhecimento é um meio natural utilizado pelos especialistas para acumular conhecimento a respeito de um determinado domínio. As regras são representadas por proposições, expressas na forma de “SE A ENTÃO C”, o antecedente “A” é formado pelo conjunto de condições e “C” representa o consequente da regra.
- *Redes Semânticas*: permitem representar o conhecimento através de modelos que são formulados como grafos (meio de se representar explicitamente relações utilizando nós e arestas), com os nós representando fatos, objetos e/ou conceitos e as arestas representando suas relações ou associações entre conceitos (LUGER, 2009).
- *Frames*: representa o conhecimento por hierarquia de classes e subclasses podendo chegar até às instâncias. Cada frame é composto por *slots* que contêm as características e propriedades da classe ou instância em questão (NIEVOLA, 1995).

2.2.3 Arquitetura de um SE

Para projetar um SE, o desenvolvedor necessita de uma estrutura básica que compõe uma arquitetura (Figura 12) capaz de armazenar o conhecimento, processá-lo e trocar mensagens com o usuário. Estas três atividades determinam de forma clara as três partes de um SE, a saber, respectivamente: base de conhecimento, mecanismo (motor/máquina) de inferência e interface com o usuário.

Três atores estão envolvidos no projeto e desenvolvimento de um SE: o especialista no domínio, engenheiro de conhecimento e o usuário (SCHULTE et al., 1987).

Especialistas são profissionais que alcançaram uma perícia em um domínio específico do conhecimento em decorrência de qualificação, aptidão ou por uma vasta experiência no desenvolvimento de procedimentos adequados à resolução de problemas do domínio em questão (TALUKDAR et al., 1986).

Engenheiro do Conhecimento é a denominação dada ao profissional que projeta SEs. Cabe a ele observar, conversar e trabalhar com o especialista humano para determinar como expressar o processo de raciocínio do especialista numa forma objetiva (MAEDCHE, 2002).

O usuário não necessariamente precisa ser um especialista, e utiliza o sistema especialista para inferir sobre questões do domínio de conhecimento, na busca por respostas em atividades específicas de um especialista.

O processo arquitetural inicia-se com a interação entre o engenheiro do conhecimento e o especialista do domínio na atividade de aquisição do conhecimento, tal conhecimento

abstraído do especialista é representado na base de conhecimento. Posteriormente, a máquina de inferência e a interface do usuário são desenvolvidas.

O usuário consulta o sistema especialista sobre problemas do domínio, o módulo de interface comunica o mecanismo de inferência que por sua vez solicita o conhecimento necessário para inferir sobre a questão. Após a inferência a interface disponibiliza respostas necessárias para auxiliar o usuário na tomada de decisão.

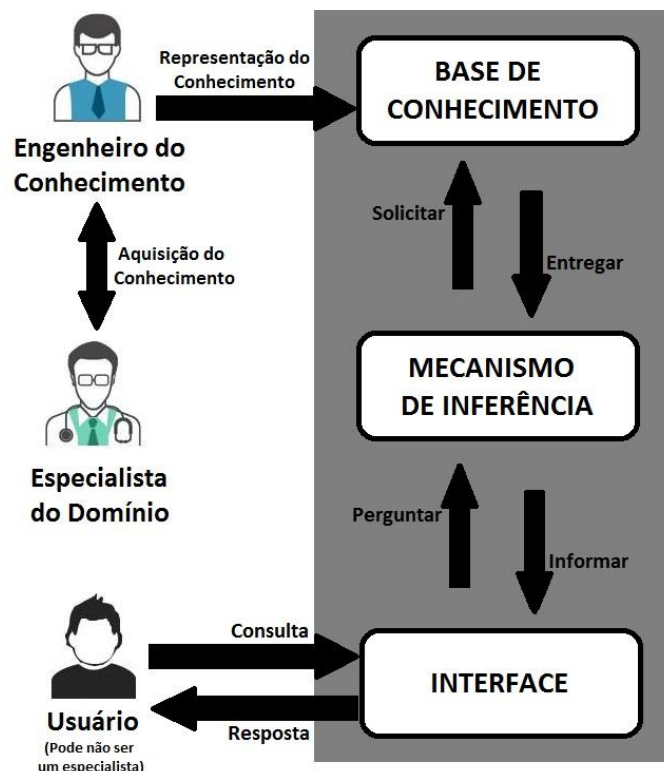


Figura 12. Arquitetura de um SE

2.2.3.1 BASE DE CONHECIMENTO

A fase de construção da base de conhecimento de um sistema especialista, frequentemente chamada de engenharia de conhecimento, é uma das mais complexas na implementação, pois o conhecimento de um especialista não se encontra formalizado, precisando portanto de um trabalho prévio para tal. A base de conhecimentos está interligada com quase todos os demais elementos do sistema, especialmente com a máquina de inferência, o mecanismo de aprendizagem e aquisição do conhecimento.

Para Genaro (1986) a base de conhecimentos de um sistema especialista compreende o conhecimento de uma área específica. O conteúdo do banco de conhecimento é essencialmente de dois tipos: conhecimento factual e conhecimento heurístico.

O conhecimento factual é representado por fatos, informações evidenciadas e aceitas pela comunidade científica, é o conhecimento contido nas publicações e livros. Por sua vez o conhecimento heurístico são as regras de “bom senso” de especialistas em alguma área da Ciência, do conhecimento dos especialistas que normalmente não têm como comprová-las cientificamente e é delas que resulta a força dos SEs (CHAIBEN, 2016).

2.2.3.2 MECANISMO DE INFERÊNCIA

O mecanismo de inferência é considerado o núcleo de um Sistema Especialista, pois é através dele que os fatos e as heurísticas contidos na base de conhecimento são aplicados no processo de solução do problema (CHAIBEN, 2016).

O processo de inferência está diretamente associado com a estrutura utilizada para o armazenamento do conhecimento. Entretanto, de forma geral, pode-se afirmar que o processo envolve um encadeamento lógico que permite tirar conclusões a partir do conhecimento existente. Conforme Heinzle (1995), “o motor de inferência é, portanto, o responsável pela ação repetitiva de buscar, analisar e gerar novos conhecimentos”.

Para Genaro (1986), a tarefa do mecanismo de inferência é selecionar e então aplicar a regra mais apropriada em cada passo da execução do sistema especialista, o que contrasta com técnicas de programação convencional, onde o programador seleciona a ordem na qual o programa deverá executar os passos, ainda em tempo de programação.

Em geral, o mesmo motor de inferências pode ser usado para fazer derivações sobre diferentes bases de conhecimento porque não contém um domínio de informação. É apenas um programa de cálculo ou de busca que analisa o conteúdo contido na base de conhecimento e aplica as regras necessárias segundo a meta estipulada pelo sistema naquele momento. A capacidade do motor de inferência é baseada em uma combinação de procedimentos de raciocínios que se processam de forma regressiva e progressiva.

Na forma de raciocínio progressivo, as informações são fornecidas ao sistema pelo usuário, que, com suas respostas, estimula o desencadeamento do processo de busca, navegando através da base de conhecimento, procurando por fatos, regras e heurísticas que melhor se aplicam a cada situação. O sistema continua nesta interação com o usuário até encontrar a solução para o problema em questão.

No modelo de raciocínio regressivo, os procedimentos de inferência acontecem de forma inversa. O sistema parte de uma opinião conclusiva sobre o assunto, podendo ser inclusive oriunda do próprio usuário, e inicia uma pesquisa pelas informações por meio das

regras e dos fatos da base de conhecimento, procurando provar se aquela conclusão é a mais adequada solução para o problema analisado.

2.2.3.3 INTERFACE DO USUÁRIO

Nos SEs a interface ganha contornos especiais, face ao caráter de documentação (citação dos textos em que se baseia a conclusão), acessibilidade (linguística e técnica) e transparência (explicitação minuciosa dos procedimentos seguidos através de mecanismos de justificação) dos conteúdos da consulta.

Para que isto ocorra, a interface deve ser flexível o bastante para que a interação entre o SE e o usuário conduza a uma eficiente navegação na base de conhecimentos durante o processamento das heurísticas, permitindo que o usuário descreva o problema ou os objetivos que deseja alcançar, e também facilita a recuperação do caminho percorrido pelo sistema para chegar à solução do problema, através de um modelo de consulta estruturado. Esse caminho é denominado *trace* e é a base de pesquisa para a explanação, que consiste em explicar “o porquê” e o “como” o sistema chegou à tal conclusão. Esse processo é muito importante, pois oferece ao usuário ajuda para julgar se adota ou não a solução apresentada pelo Sistema Especialista (MENDES, 2016)

2.3 Reconhecimento de Padrões

Segundo Pao (1989), o conhecimento que envolve o reconhecimento de padrões faz parte do cotidiano do ser humano. A formação da linguagem, o modo de falar, o desenho das figuras, o entendimento das imagens, tudo envolve padrões. Reconhecimento de padrões é uma tarefa complexa, onde o homem busca, sempre, avaliar as situações em termos dos padrões das circunstâncias que as constituem, descobrir relações existentes no meio para melhor entendê-lo e adaptar-se a ele.

Tal habilidade advém de um longo processo evolutivo, assim como outros sentidos inerentes ao homem como percepção de espaço, localização e distância, situadas como determinante para a sobrevivência da espécie. No entanto, a capacidade de reconhecimento de padrões do homem apresenta limitações com o aumento da complexidade, nomeadamente na identificação de características e propriedades dos objetos, no volume de informações, ou na carga de operações matemáticas e lógica.

Desta forma, os sistemas computacionais buscam suprir as limitações humanas com o aumento da perícia no reconhecimento de padrões em aplicações cada vez mais desafiadoras,

o que requer uma maior exigência computacional, promovendo a contribuição das mais diversas áreas de pesquisas distintas (e.g. sistemas e processamento de imagens e sinais, inteligência artificial, modelagem conexionista, teoria de estimação/otimização, conjuntos difusos, modelagem estrutural, linguagem formal, algoritmos de classificação e clusterização). Isto é, por si só, uma indicação do sucesso da extensão e profundidade de interesse no tópico e do vigor das pesquisas associadas.

Para auxiliar na compreensão sobre o significado de “padrão”, a Tabela 2 apresenta de modo resumido algumas definições de autoria dos relevantes estudiosos da área. Na sequência, apresenta-se a Tabela 3 com as definições de Reconhecimento de Padrões utilizadas na literatura contemporânea.

Tabela 2. Definição de Padrão

AUTOR	DEFINIÇÃO
(TOU e GONZÁLES, 1981)	Padrões são propriedades ou características que definem um objeto ou um grupo de objetos que possibilitem o seu agrupamento, arranjo ou estrutura organizacional entre os objetos semelhantes dentro de uma determinada classe ou categoria, mediante a interpretação de dados de entrada, que permitam a extração das características relevantes desses objetos
(JAIN et al., 2000)	Um padrão é algo que segue alguma regra ou conjunto de regras, de forma que seja possível distingui-lo de outros padrões. Por sua vez, o reconhecimento de padrões é a capacidade de reconhecer e de diferenciar os diversos padrões existentes
(FU e MUI, 1981)	Um padrão é uma estrutura de medidas quantitativas e qualitativas que representa alguma entidade na imagem origem. Um descritor é uma das medidas que compõem a estrutura do padrão. Em geral um padrão é formado de vários descritores, na quantidade necessária para classificá-lo, arranjado de forma a fornecer informações adequadas a respeito do padrão em questão.
(CORDEIRO, 2002)	Tratando-se de imagens, um padrão é o conjunto de medidas quantitativas e qualitativas que representam alguma entidade na imagem de origem, extraída no processo de identificação do padrão. Intensidade de sinais, cores, tons de cinza e geometrias,

	são exemplos de medidas. Estas medidas podem ser simbólicas, numéricas ou ambas, geralmente sendo representados em forma de vetor ou matriz.
--	--

A complexidade do reconhecimento de padrões está na estrutura do padrão, na escolha das características e das propriedades que irão definir o padrão, o que pode não ser uma tarefa trivial. Deve-se escolher e extrair um conjunto finito de características que represente totalmente o padrão em questão e que seja passível de ser manuseado (BEZDEK e PAL, 1992).

Tabela 3. Reconhecimento de Padrões (RP)

AUTOR	DEFINIÇÃO
(DUDA e HART, 1973)	<i>Campo que consiste no reconhecimento de regularidades significativas em meios ruidosos e complexos</i>
(BEZDEK e PAL, 1992)	A busca por estruturas em dados
(JAIN et al., 2000)	Um estudo de como máquinas podem observar o ambiente, aprender a distinguir padrões de interesse do seu propósito, e tomar decisões concretas e aproximadas sobre suas categorias
(TOU e GONZÁLES, 1974)	Reconhecimento de padrões é o processo de identificar objetos, através da extração de suas características, a partir de dados sobre o objeto

Um método de reconhecimento de padrões analisa as características de um elemento e classifica-o em um grupo pré-definido contido na base de conhecimento. Problemas de reconhecimento de dígitos, reconhecimento de faces, predição de tendências em séries financeiras, predição de falha sem equipamentos e muitos outros, englobam o universo do reconhecimento de padrões.

2.3.1 Reconhecimento de padrões de imagem

O trabalho de reconhecimento de padrões pode ser dividido em duas etapas: o pré-processamento e o reconhecimento propriamente dito (BISHOP, 1995). No pré-processamento são retiradas características do objeto a ser reconhecido e estas características

serão utilizadas para facilitar o trabalho da classificação, eliminando informações que não sejam úteis e possam vir a atrapalhar o trabalho de reconhecimento.

Para imagens digitais, o pré-processamento também é dividido em duas fases distintas: a segmentação da imagem e a extração das características que serão analisadas como mostrado na Figura 13. Na segmentação da imagem, o objeto a ser reconhecido é isolado do resto da imagem. Na extração das características, serão selecionados atributos significativos da imagem, formando um vetor de atributos, de forma que este possa representar simplificada a imagem, diminuindo a quantidade de informação necessária para classificá-la, e, conseqüentemente, o tempo de processamento para executar a tarefa (CASTLEMAN, 1996).

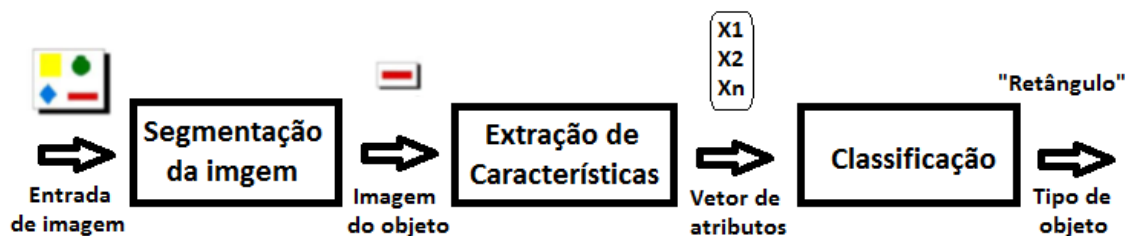


Figura 13. As três fases do reconhecimento de padrões de imagens digitais

Fonte: (CASTLEMAN, 1996).

Deve ser ressaltado que não existe um modelo formal para aplicação das etapas do processo. A segmentação e a extração de características são processos empíricos e adaptativos, procurando sempre se adequar às características particulares de cada tipo de imagem e aos objetivos a serem alcançados por meio de técnicas de segmentação e de extração de objetos específicos ao contexto, como: **Histograma**, **GLCM**(do inglês, *Gray Level Co-occurrence Matrix*), **Descritores de Haralick** (HARALICK, 1973)(HARALICK, 1979), **Filtro de Gabor** (GABOR, 1946)(LEE e WANG, 1999)(MARCELJA, 1980)(DAUGMAN, 1985)(FOGEL e SAGI 1989), **Transformada de Fourier** (GONZALEZ e WOOD, 2002), **Transformada de Wavelets** (GONZALEZ e WOOD, 2002), **PCA** (do Inglês, *Principal Components Analysis*) (LUDWIG e REYNOLDS, 1988) (ODDEN e KVALHEIM, 2000)(KVALHEIM, 1998), **Fractais** (BARNESLEY e DEMKO, 1985)(JINJIANG, DA YUAN e ZHANG, 2008), **LBP** (do inglês, *Local Binary Pattern*), **LPQ** (do inglês, *Local Phase Quantization*) (OJALA e PIETIKÄINEN, 1996) (OJANSIVU E HEIKKILÄ, 2008)(JUN e KIM, 2011), **Filtro Passa-Alta e Filtro Passa-Baixa** (GONZALEZ e WOOD, 2002).

2.3.2 Segmentação: Histograma

Segmentar uma imagem significa, de modo simplificado, separar a imagem como um todo nas partes que a constituem e que se diferenciam entre si. É usual denominar *objetos* da imagem os grupos de pixels de interesse, ou que fornecem alguma informação para o Processamento Digital de Imagem (PDI). Da mesma forma, a denominação *fundo* da imagem é utilizada para o grupo de pixels que podem ser desprezados ou que não têm utilidade no processamento de imagem. Essas denominações (*objeto* e *fundo*) possuem uma conotação bastante subjetiva, podendo se referir a grupos de pixels que formam determinadas regiões na imagem sem que representem um objeto, de modo literal, presente na imagem processada (GAGVANI, 2008).

A segmentação é considerada, entre todas as etapas do processamento de imagens, a etapa mais crítica do tratamento da informação. É na etapa de segmentação que são definidas as regiões de interesse para processamento e análise posteriores. Como consequência deste fato, quaisquer erros ou distorções presentes nesta etapa se refletem nas demais etapas, podendo produzir ao final do processo, resultados não desejados que possam contribuir de forma negativa para a eficiência de todo o processamento.

O histograma de uma imagem, parte integrante da segmentação, é um conjunto de números indicando o percentual de pixels naquela imagem, que apresentam um determinado nível de cinza (MARQUES, 1999). Através da visualização do histograma de uma imagem obtemos uma indicação de sua qualidade quanto ao nível de contraste e quanto ao seu brilho médio (se a imagem é predominantemente clara ou escura).

Os histogramas são ferramentas de processamento de imagens que possuem grande aplicação prática. São frequentemente usados em qualquer estudo para representar uma grande quantidade de dados numéricos, como meio para analisar as informações de forma mais fácil e simples, do que por meio de uma grande tabela (MENESES e ALMEIDA, 2012). Segundo Marengoni e Stringhini (2009) os histogramas são determinados a partir de valores de intensidade dos pixels. Entre as principais aplicações dos histogramas estão a melhora da definição de uma imagem, a compressão de imagens, a segmentação de imagens ou ainda a descrição de uma imagem. O histograma de uma imagem I , cujos valores de intensidade estejam entre 0 e G (máximo valor de pixels), é definido pela equação 1:

$$h(I_k) = n_k$$

Equação 1

onde I_k é um valor de intensidade k , ($0 \leq k \leq G$) da imagem I e n_k é o número de pixels na imagem I que possuem a intensidade k . É possível normalizar um histograma, representando os valores em termos de porcentagem, conforme mostrado na equação 2:

$$p(I_k) = \frac{h(I_k)}{n} = \frac{n_k}{n} \quad \text{Equação 2}$$

onde n é o número de pixels da imagem. A Figura 13 mostra como um histograma é determinado.

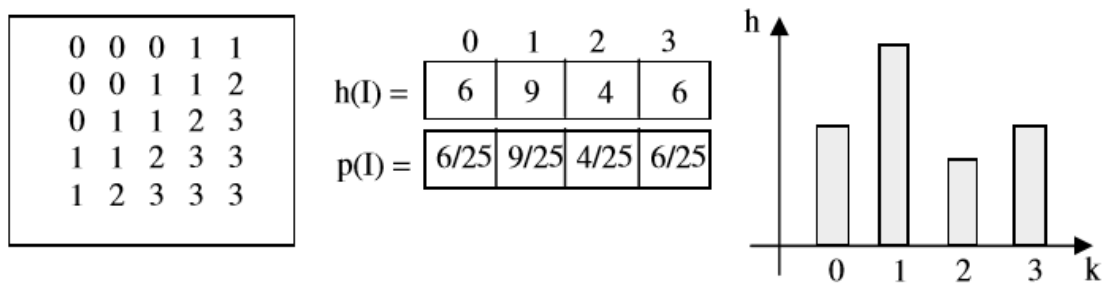


Figura 14. À esquerda uma imagem I , ao centro o histograma da imagem em valores ($h(I)$) e em porcentagem ($p(I)$), à direita uma representação do histograma de forma gráfica.

Uma operação bastante comum utilizando histogramas é o ajuste dos valores de intensidade de forma a melhorar o contraste em uma imagem. Esta operação é chamada de equalização de histogramas. Esta operação mapeia os valores de intensidade de uma imagem de um intervalo pequeno (pouco contraste) para um intervalo maior (muito contraste) e ainda distribui os pixels ao longo da imagem, a fim de obter uma distribuição uniforme de intensidades (embora na prática isso quase sempre não ocorra) (MARENGONI e STRINGHINI, 2016). A expressão que fornece um histograma equalizado é apresentada na equação 3:

$$h_{eq}(k) = \frac{(L-1)}{MN} \sum_{j=0}^k n_j \quad \text{Equação 3}$$

onde k é a intensidade no histograma equalizado, L é o valor máximo de intensidade na imagem, M e N são as dimensões da imagem e n_j é o número de pixel na imagem com valor de intensidade igual a j . A Figura 15 mostra um exemplo de uma imagem que foi ajustada utilizando equalização de histograma.

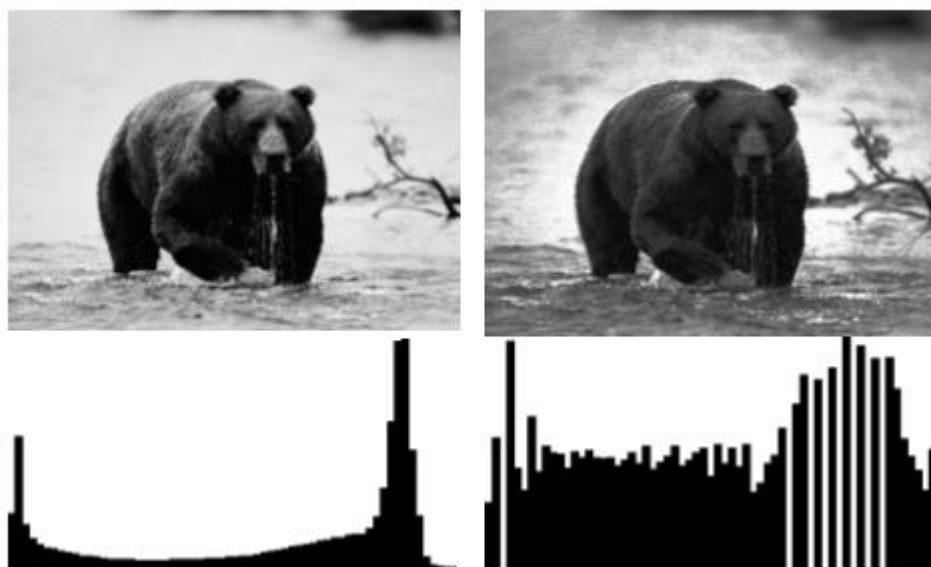


Figura 15. Topo à esquerda, imagem em nível de cinza, e abaixo dela o histograma da imagem. Topo à direita, a mesma imagem após equalização, e o histograma equalizado da imagem.

2.3.3 Extração de Características: Descritores de Haralick

Muitas vezes o resultado da segmentação não é adequado para que os grupos de pixels segmentados sejam representados e descritos em termo de suas características nas etapas subsequentes. Sendo assim, são necessárias técnicas de extração de características representativas para posterior classificação.

A extração de características é onde se inicia a etapa propriamente dita de análise da imagem. Nesta etapa são realizadas medidas na imagem segmentada ou pós-processada, ou até mesmo na imagem em tons de cinza. Através dessas medidas, os grupos de pixels são descritos por atributos característicos, gerando dados quantitativos para o objetivo final.

Desta forma, na etapa de extração de características Haralick e colaboradores (1973) descrevem uma metodologia para descrição da textura da imagem, onde são definidas diversas características advindas do cálculo de matrizes de coocorrência, que são matrizes que contam as ocorrências de níveis de cinza em uma imagem. Essas características servem como medida para a diferenciação de texturas que não seguem um determinado padrão de repetitividade, fornecendo informações relevantes para classificação das mesmas, como observado em Haralick, Shanmugam e Dinstein (1973).

Algumas características de Haralick são descritas a seguir, em termos de sua significância (HARALICK et al, 1973)(ROSENFELD e KAK, 1982)(AKSOY e HARALICK, 1999) (PEREZ, GONZAGA e ALVES, 2001):

- *Segundo Momento Angular (SMA)*: medida da homogeneidade local dos níveis de cinza em uma imagem. Em uma imagem homogênea existem poucas transições de níveis de cinza. Nesse caso, a matriz de coocorrência possui baixas entradas de alta magnitude.

$$SMA = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} [P(i, j, d, \theta)]^2 \quad \text{Equação 4}$$

- *Contraste ou Variância*: medida da quantidade de variação local de níveis de cinza em uma imagem. Rosenfeld e Kak (1982) afirmam que se este valor for pequeno, os níveis de cinza de imagem analisada são todos próximos de uma média, ou seja, a matriz de coocorrência de níveis de cinza vai possuir a maioria dos valores concentrados na horizontal. Caso contrário, se o valor do contraste for alto, a imagem possui uma maior distribuição dos níveis de cinza em seu histograma.

Equação 5

$$\text{Contraste} = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} (i - j)^2 P(i, j, d, \theta)$$

- *Entropia*: A Entropia ou grau de dispersão de níveis de cinza pode também, juntamente com o SMA, ser utilizada como medida da homogeneidade em uma imagem. O valor da Entropia é alto quando os valores da matriz de coocorrência são iguais e é baixo quando a concentração de valores na diagonal é alta.

$$\text{Entropia} = - \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} P(i, j, d, \theta) \log_2 [P(i, j, d, \theta)] \quad \text{Equação 6}$$

- *Momento Diferença Inverso (MDI)*: Segundo Rosenfeld e Kak (1982), o MDI atinge seu valor máximo quando a concentração dos valores na diagonal da matriz de coocorrência for máxima.

$$MDI = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \frac{1}{1 + (i - j)^2} P(i, j, d, \theta) \quad \text{Equação 7}$$

- *Correlação*: A Correlação representa uma idéia de linearidade de dependências de tons de cinza em uma imagem. Conforme Haralick, Shanmugam e Dinstein (1973), em uma imagem onde exista certa ordenação local de níveis de cinza, o valor da correlação é alto.

$$\text{Correlação} = \frac{\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} ij P(i, j, d, \theta) - \mu_x \mu_y}{\sigma_x \sigma_y}, \text{ sendo}$$

Equação 8

Existem diversas outras características de Haralick utilizadas para extração de características em texturas para diferenciação de imagens, conforme Haralick e colaboradores (1973). As características discutidas acima foram escolhidas de acordo com pesquisa bibliográfica, onde, a partir de diversos trabalhos usando a implementação dessas características, foram, então, selecionadas aquelas que possam levar a resultados mais relevantes (WESZKA et al, 1976)(CONNERS et al., 1980)(CONNERS, TRIVEDI et al., 1984).

2.3.4 Classificação: Redes Neurais Artificiais

Classificar padrões em uma imagem é um dos processos mais complexos em processamentos de imagens digitais. Abordagens diferentes são utilizadas visando diminuir o custo computacional e que permitam utilizar os dados extraídos em diferentes domínios de aplicações (PEDRINI e SCHWARTZ, 2008).

O diagnóstico por imagem é uma importante técnica de investigação. As informações extraídas dos padrões existentes em uma imagem podem auxiliar a tomada de decisões em várias áreas, tais como o sensoriamento remoto, medicina, recuperação de imagens, controle de qualidade e em microscopia (NASCIMENTO et al., 2003). Dessa forma, o adequado ajuste das técnicas de melhorias e sua classificação são de fundamental importância para que as imagens obtidas se tornem um instrumento auxiliar significativo para a tomada de decisões em áreas como as supracitadas.

Entre as técnicas computacionais que permitam a classificação de padrões em uma imagem, destacam-se os seguintes classificadores: **RNA** (Rede Neural Artificial)(HAYKIN, 2009), **SVM** (do inglês. *Support Vector Machine*)(MUKHERJEE et al., 1999)(BROWN et al., 2000), **KNN** (do inglês. *K-Nearest Neighbors*) (BEYER e GOLDSTEIN, 1999)(GOLDSTEIN e RAMAKRISHNAN, 2000)(KATAMAYA e SATOH, 1997), **Algoritmo de Otsu** (OTSU, 1975)(LIAO, CHEN e CHUNG, 2001), **Algoritmo de Cross-bin** (MA, GU e WANG, 2010)(KURTZ et al., 2013) e **Erro Médio Quadrático** (SCHERB, KUEHN e KAMMEYER, 2002)

As RNAs, também conhecidas como métodos conexionistas, são inspiradas nos estudos da maneira de como se organiza e como funciona o cérebro humano. Trata-se de modelos computacionais não lineares, inspirados na estrutura e na operação do cérebro humano, que procuram reproduzir características humanas, tais como: *aprendizado, associação, generalização e abstração*. As Redes Neurais são efetivas no aprendizado de padrões a partir de dados não lineares, incompletos, com ruído ou compostos de exemplos contraditórios (PACHECO, 2002).

Uma RNA é um sistema que tem capacidade computacional adquirida por meio de aprendizado e generalização (BRAGA, CARVALHO e LUDEMIR, 2003). O aprendizado está relacionado com a capacidade das RNAs de adaptar seus parâmetros como consequência com a interação com o ambiente externo. A generalização, por sua vez, está associada à capacidade destas redes de fornecerem respostas consistentes para dados não apresentados durante a etapa de treinamento.

2.3.4.1 APRENDIZADO CONEXIONISTA

O aprendizado ocorre quando a rede neural atinge uma solução generalizada para uma classe de problemas. Denomina-se algoritmo de aprendizado a um conjunto de regras bem definidas para a solução de um problema de aprendizado. Existem muitos tipos de algoritmos que diferem entre si, principalmente, pelo modo como os pesos⁶ são modificados. Aprendizagem, para uma rede neural, envolve o ajuste destes pesos (DHAR e STEIN, 1997).

A RNA baseia-se nos dados para extrair um modelo geral. Portanto, a fase de aprendizado deve ser rigorosa e verdadeira, a fim de se evitar modelos espúrios.

Todo o conhecimento de uma rede neural está armazenado nas sinapses, ou seja, nos pesos atribuídos às conexões entre os neurônios. De 50% a 90% do total de dados devem ser separados para o treinamento da rede neural, dados estes escolhidos de forma criteriosa buscando contemplar as diferentes particularidades dos domínios das variáveis envolvidas, a fim de que a rede seja capaz de “aprender” e “generalizar”. O restante dos dados só é apresentado à rede neural na fase de validação.

O aprendizado conexionista, de modo geral, é um processo gradual e interativo, onde os pesos são modificados várias vezes, pouco a pouco, seguindo-se uma regra de aprendizado que estabelece a forma como estes pesos são alterados. O aprendizado é realizado utilizando-se um conjunto de dados de aprendizado disponível (base de exemplos). Cada interação deste

⁶ Peso: é a intensidade da força sináptica e pode ser fixo ou treinável implementando as ligações entre as unidade e a intensidade com que o sinal é transmitido de um neurônio ao outro (DAYHOFF, 1992).

processo gradativo de adaptação dos pesos de uma rede neural, sendo feita uma apresentação completa do conjunto de dados, é chamada de *época de aprendizado*. Os métodos de reconhecimento de padrões (métodos de aprendizado neural) podem ser agrupados em duas grandes categorias: supervisionada e não-supervisionada:

- **Aprendizado supervisionado:** o usuário dispõe de um comportamento de referência preciso que ele deseja ensinar à rede. Sendo assim, a rede deve ser capaz de medir a diferença entre seu comportamento atual e o comportamento de referência e, então, corrigir os pesos de maneira a reduzir este erro (desvio de comportamento em relação aos exemplos de referência). Exemplo de aplicação: reconhecimento de caracteres em uma aplicação do tipo OCR (*Optical Character Recognition*) (OSÓRIO, 1991).
- **Aprendizado não-supervisionado:** os pesos da rede são modificados em função de critérios internos, tais como, por exemplo, a repetição de padrões de ativação em paralelo de vários neurônios. O comportamento resultante deste tipo de aprendizado é usualmente comparado com técnicas de análise de dados empregadas na estatística (e.g. *clustering*). Exemplo de aplicação: diferenciar tomates de laranjas, sem no entanto ter os exemplos com a sua respectiva classe etiquetada (e.g. *self-organizing feature maps*) (KOHONEN, 1987).

2.3.4.2 MULTILAYER PERCEPTRON

O primeiro modelo de rede neural implementado foi a *perceptron*, por Frank Rosenblatt em 1958. A RNA *perceptron* contém uma camada de entrada e uma de saída, limitando-se a fronteiras de decisão linear e funções lógicas simples.

Minsky e Papert (1969) analisaram matematicamente o *perceptron* e demonstraram que redes de uma só camada não são capazes de solucionar problemas que não sejam linearmente separáveis. Como não acreditavam na possibilidade de se construir um método de treinamento para redes com mais de uma camada, eles concluíram que as redes neurais seriam sempre suscetíveis a essa limitação (RUSSELL e NORVIG, 2004).

Identificados as limitações relativas ao *perceptron* de camada simples, foi desenvolvida a RNA *multilayer perceptron* que, em geral, consiste de uma camada de entrada, uma ou mais camadas intermediárias (escondidas) e uma camada de saída. Trata-se de um caso em particular de topologia de rede, suprindo as limitações de linearidades como se pode observar na Figura 16 (VELLASCO, 2000).

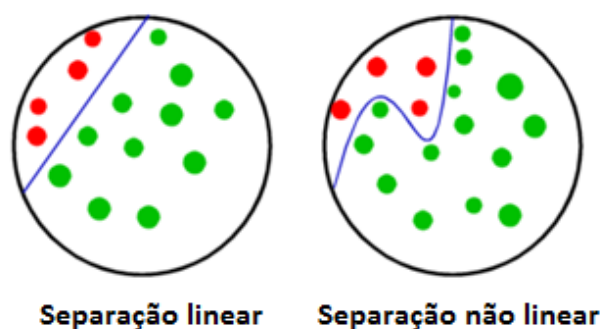


Figura 16. Linearidade das redes neurais (KROGH e VEDELSBY, 1995)

Nas RNAs *multilayer perceptron*, cada camada tem uma função específica. A camada de saída recebe os estímulos da camada intermediária e constrói o padrão que será resposta. As camadas intermediárias funcionam como extratoras de características, seus pesos são uma codificação de características apresentadas nos padrões de entrada e permitem que a rede crie sua própria representação, mais rica e complexa do problema.

2.3.4.3 ALGORITMO BACKPROPAGATION

O algoritmo Backpropagation e suas derivações é um dos algoritmos para treinamento de RNAs multicamadas, mais difundido na comunidade científica. Baseia-se no aprendizado supervisionado por correção de erros. Sua utilização compreende duas fases de propagação (Figura 17)(Figura 18), descritas a seguir.

1º - Propagação: Um padrão é apresentado à camada da entrada da rede. A atividade resultante flui através da rede, camada por camada, até que a resposta seja produzida pela camada de saída, onde é obtido a resposta da rede e o erro é calculado (CARVALHO, 1998).

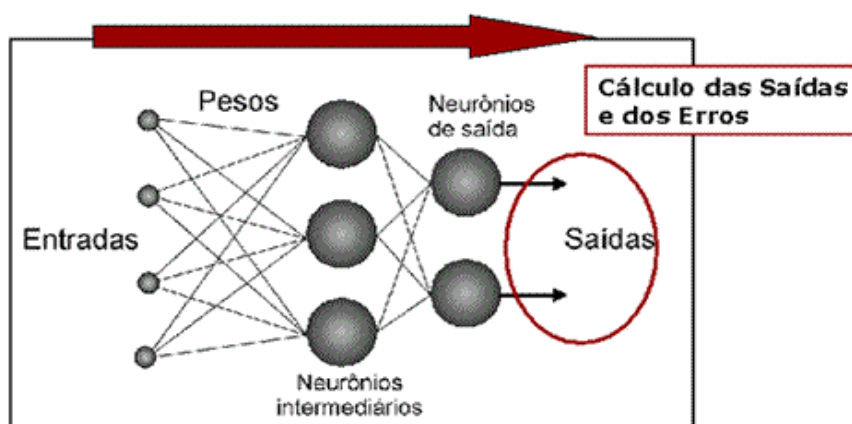


Figura 17. Fase de Propagação (LNCC 2008)

2º - Retropropagação ("backpropagation"): Desde a camada de saída até a camada de entrada, são feitas alterações nos pesos sinápticos.

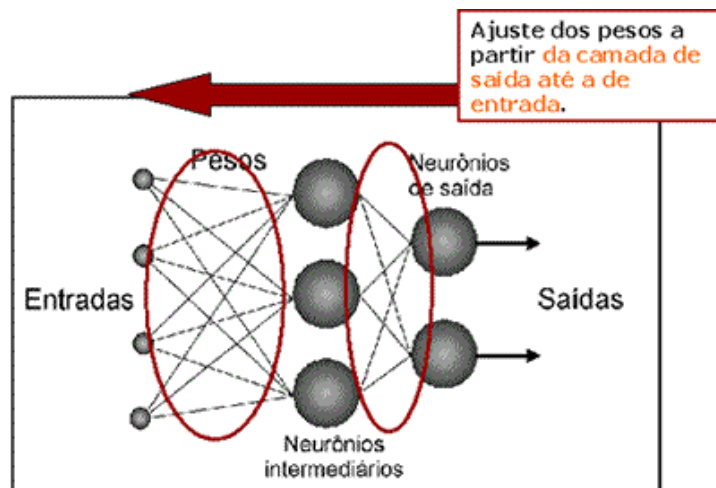


Figura 18. Fase de Retropropagação (LNCC 2008)

Depois que a rede estiver treinada e o erro estiver em um nível satisfatório, ela poderá ser utilizada como uma ferramenta para classificação de novos dados. Para isto, a rede deverá ser utilizada apenas no modo progressivo (*feed-forward*). Ou seja, novas entradas são apresentadas à camada de entrada, são processadas nas camadas intermediárias e os resultados são apresentados na camada de saída, como no treinamento, mas sem a retropropagação do erro. A saída apresentada é o modelo dos dados, na interpretação da rede (CARVALHO, 1998).



3. TRABALHOS CORRELATOS

Este capítulo apresenta as principais contribuições científicas relacionadas ao tema central desta pesquisa, levando em consideração os mais relevantes e referidos na literatura. Os trabalhos estão divididos em duas grandes áreas: Reconhecimento de Padrões (especificando trabalhos que envolvem imagem botânica) e Ontologias (categorizadas em domínio florestal, monitoramento e impacto florestal, ecossistema e biodiversidade).

3.1 Reconhecimento de Padrão de Imagem Botânica

O reconhecimento de padrões é definido como o processo pelo qual um padrão recebido é atribuído a uma classe dentre um número pré-determinado de classes (HAYKIN, 2009). Assim sendo, a Figura 19 ilustra o processo de reconhecimento de padrões de imagens de madeira utilizado neste trabalho para representar a organização dos trabalhos relacionados com esta pesquisa. O processo é dividido em 3 etapas, nomeadamente captura da amostra, segmentação e classificação.

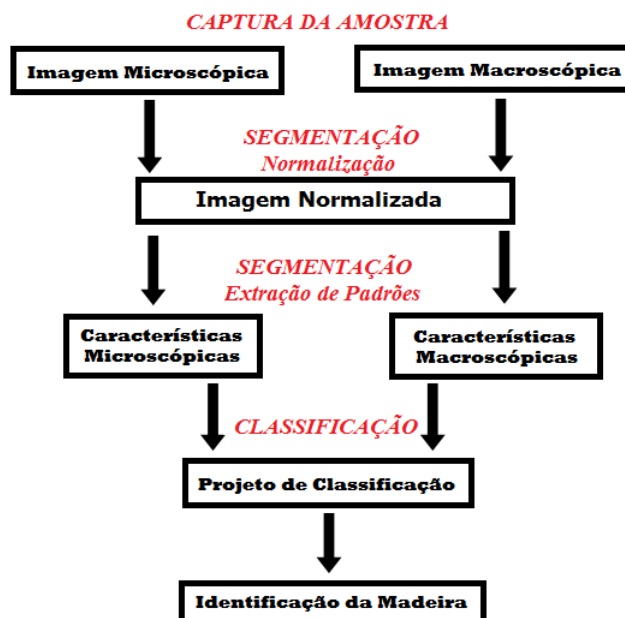


Figura 19. Processo de reconhecimento de padrão

A captura da amostra ocorre por equipamento com função de digitalização de imagens com qualidades de resoluções e aproximação diferentes, que visam contemplar a necessidade e finalidade da imagem. As imagens dividem-se em macroscópicas e microscópicas. Dentro do campo da microscopia são utilizadas imagens por microscópios padrões (imagens por Stereograma) e imagem com microscópio eletrônico de varredura (Figura 20).

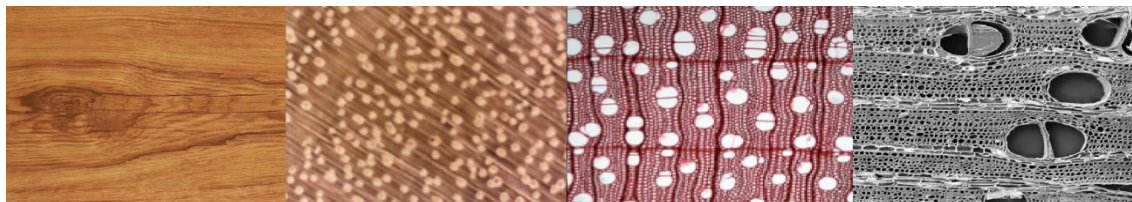


Figura 20. Imagens Madeira

A etapa da segmentação subdivide-se em normalização e extração de padrões. Na normalização ocorre uma preparação da imagem a ser processada, permitindo a padronização do campo de amostragem definindo tamanho, cores, formatos e aproximação da amostra. Na extração de padrões ocorre o pré-processamento e são retiradas características do objeto a serem reconhecidas. Estas características serão utilizadas para facilitar o trabalho da classificação, eliminando informações que não sejam úteis e possam vir a atrapalhar o trabalho de reconhecimento.

A classificação é a parte mais abstrata do processo de visão computacional. Nesta etapa ocorre o reconhecimento, o que permite obter a compreensão e a descrição final da imagem analisada. A classificação parte da premissa que a similaridade entre objetos implica que eles possuam características similares, formando classes. O resultado da classificação pode ser percentual (indicando % de chance da ocorrência de alguma classe) ou também pode ser uma imagem com algumas características enfatizadas para auxiliar o especialista em sua tomada de decisão. De outra maneira, podemos considerar que a fase de classificação consiste em reconhecer um objeto, uma forma ou, de modo geral, uma entidade particular da imagem. Dado um conjunto de classes e um padrão apresentado como entrada para o sistema. O problema consiste em decidir a que classe o padrão pertence (WHELAN e MOLLOY, 2001)(GONZALEZ e WOOD, 2002).

A descrição dos trabalhos correlatos segue a organização apresentada no processo de reconhecimento de padrões de imagens de madeira, apresentando como os trabalhos relevantes neste campo da pesquisa capturam as amostras, segmentam e classificam as imagens da madeira.

3.1.1 Captura da Amostra

O processo de aquisição de imagem repercute em toda estrutura do sistema de reconhecimento de padrão, posto que a imagem como foco principal do sistema apresenta características que variam conforme a tecnologia empregada para sua captura. Neste sentido, observa-se na Tabela 4 os principais trabalhos relacionados a esta pesquisa e seus respectivos processos de captura da amostragem de imagem da madeira.

Tabela 4. Imagens

Autor	Imagem Macroscópica	Imagem Microscópica		
		Microscópio convencional	Stereograma	Microscópio eletrônico de varredura
(XUEBING, 2005)	X			
(BIHUI et al., 2010)			X	
(DE PAULA e TUSSET, 2009)	X			
(DE PAULA FILHO et al., 2014)	X			
(DE PAULA FILHO, 2012)	X			
(HANGJUN et al., 2009)		X		
(HANGJUN et al., 2011)		X		
(HANGJUN et al., 2012-1)		X		
(HANGJUN et al., 2012-2)			X	
(HENGNIAN et al., 2008)		X		
(KHALID, 2008)	X			
(LINGJUN et al., 2011-1)			X	
(LINGJUN et al., 2011-2)			X	
(LINJIN et al., 2012)			X	
(MALLIK et al., 2011)				X
(MARTINS et al., 2012)		X		
(TOU, 2007)	X			
(HUI, 2009)	X			
(SHAOCHUN, 2007)	X			

(HAIPENG, 2007)	X			
(ZHIWEI et al., 2011)		X		

Xuebing (2005) utiliza imagens macroscópicas de madeira para classificar espécies botânicas. Sobre a textura da superfície da imagem aplica-se o método da Matrix de Co-ocorrência de Níveis de Cinza (*Gray Level Co-occurrence Matrix* - GLCM). Neste método, os parâmetros obtidos com a Matriz de Co-ocorrência (e.g. contraste, correlação, entropia, soma dos quadrados e momento da diferença inverso) foram aplicados à textura e selecionados por análise de relevância. Os resultados mostraram que os parâmetros da Matriz de co-ocorrência são adequados para descrever a textura de madeira e desta forma identificar espécies botânicas.

Na perspectiva de amostragem de imagens microscópicas, Hangjun e colaboradores (2012-1) utilizam como ferramenta de captura de imagem um microscópio convencional com aproximação de 500 vezes. De posse da imagem microscópica da madeira, desenvolvem uma segmentação utilizando conjuntos de níveis com intuito de diminuir ruídos da madeira como bolhas, seiva e outras características que causam inconsistência na imagem da madeira, buscando a tentativa da homogeneidade dos componentes invariáveis existentes na madeira.

Sun Lingjun e colaboradores (2011-1) apresentam um reconhecimento automático de madeira através do equipamento de Estereograma, onde se atribui como pré-processamento a normalização da imagem da madeira, utilizando os fatores de Padrão do Local Binário (*Local Binary Pattern*, LBP), usado para descrever a característica da textura local de imagem, baseado nas características da madeira extraída, e na classificação utilizando o algoritmo *K-Nearest Neighbor*⁷ (KNN). A melhor taxa de reconhecimento é superior a 93%.

Mallik e colaboradores (2011) classificam as espécies de madeira por meio de microscopia eletrônica de varredura, usando imagens obtidas com ampliação de 1500 vezes com um processamento por segmentação destas imagens. Os resultados mostraram que a imagem extraída da microscopia apresenta nítida diferença na textura entre as espécies de madeira. As micrografias obtidas com o microscópio foram tratadas de forma simples, usando segmentação que contém *thresholding*⁸, borda detecção⁹ e reconhecimento de objetos para identificar as traqueídes pertencentes ao lenho inicial de sete espécies botânicas. Em seguida,

⁷ Algoritmo KNN: é uma técnica amplamente empregada para reconhecer padrões. O centro de seu funcionamento está em descobrir o vizinho mais próximo de uma dada instância.

⁸ *Thresholding*: método de segmentação que tem como função transformar uma imagem de tons de cinza se torne numa imagem binária (GONZALEZ & WOODS, 2002).

⁹ borda detecção: a abordagem mais comum para detecção de discontinuidades. Em uma imagem a borda é definida como sendo o limite entre duas regiões com diferentes propriedades. (GONZALES E WOODS, 1987).

a forma, o número e a distribuição das traqueídes foram analisados em cinco características: circularidade, retangularidade, número de traqueídeos, distância entre traqueídeos e área média. Tais características permitem a análise e a identificação das espécies.

3.1.2 Segmentação

A segmentação da imagem é a etapa mais delicada do processamento da imagem digital, tendo em vista que todo trabalho posterior será baseado na imagem segmentada. Existem inúmeras técnicas utilizadas na segmentação da imagem, e cada técnica é definida de acordo com o problema a ser resolvido, não existindo um modelo formal para o processo, que deverá se ajustar de acordo com o tipo de imagem estudada.

A Tabela 5 apresenta os trabalhos correlatos com esta pesquisa, identificando suas respectivas técnicas de segmentação aplicadas a imagens de madeira. Identifica-se que em alguns trabalhos duas ou mais técnicas de segmentação foram empregadas para melhorias das taxas de reconhecimento de padrão.

Tabela 5. Descritores de textura

Autor	Cor / Histograma	Textura	GLCM	Gabor wavelets	Haralick	PCA	Fractais	LBP	LPQ	Filtro passa- alta
(XUEBING, 2005)			X							
(BIHUI et al., 2010)			X		X					
(DE PAULA FILHO et al., 2014)	X	X	X				X			
(DE PAULA FILHO, 2012)	X	X	X	X			X	X	X	
(HANGJUN et al., 2009)	X									
(HANGJUN et al., 2011)						X				
(HANGJUN et al., 2012-1)		X								
(HANGJUN et al., 2012-2)		X		X	X					
(HENGNIAN et		X								

al., 2008)										
(KHALID, 2008)	X		X							X
(LINJIN et al., 2012)		X								
(MALLIK et al., 2011)		X								
(MARTINS et al., 2012)								X	X	
(PAULA E TUSSET, 2009)	X									
(LINGJUN et al., 2011-1)								X		
(LINGJUN et al., 2011-2)				X						
(TOU, 2007)	X				X					X
(HUI, 2009)	X									
(SHAOCHUN, 2007)	X									
(HAIPENG, 2007)		X								
(ZHIWEI et al., 2011)		X								

O histograma de uma imagem descreve a distribuição estatística dos níveis de cinza (GONZALEZ e WOODS, 2002), ou seja, mostra a frequência com que cada nível de cinza aparece na imagem. O histograma, juntamente com os níveis de coloração da imagem, representam ferramentas simples e extremamente úteis no atual cenário da segmentação. Como exemplo, Paula e Tusset (2009) utilizam-se de uma abordagem de baixo custo computacional com imagem não tratada de amostras de madeira da flora brasileira. Extraem-se características de informações de coloração, enfocando canais de cor e histograma. Tais características integram vetores usados para o reconhecimento de padrões, usando uma RNA buscando-se assim similaridades desta amostra com outras amostras provenientes de uma base de treinamento. Foram usadas para este estudo um total de 163 imagens de 14 espécies florestais distintas. As taxas de reconhecimento chegaram a 80,9% em espécies limitadas.

Os padrões de textura encontrados em imagens possuem informações sobre a distribuição espacial, luminosidade e arranjo estrutural da superfície em relação às regiões

vizinhas (HARALICK, 1979). A representatividade da textura como premissa para os métodos de segmentação dos sistemas de reconhecimento de padrão de imagem é evidenciada pelo extenso volume de aplicações no cenário científico e mercadológico. A exemplo disto, Linjin e colaboradores (2012) elaboraram um método para o reconhecimento de madeira com base em análise de textura. Com pré-processamento, as imagens de textura de madeira foram divididas em vários blocos. Posteriormente, as características da madeira foram extraídas dessas imagens bloqueadas em escala de cinza, utilizando a técnica aprimorada da Auto Correlação Local de ordem superior.

Os Descritores de Haralick (HARALICK et al., 1973) descrevem um método de classificação baseado em texturas realizado a partir de cálculos estatísticos de segunda ordem (relação entre dois pixels, o de referência e os vizinhos), que definem diversas características obtidas através de métodos de segmentação como matrizes de coocorrência, Filtro de Garbor e Filtro Passa Alta. Essas características são utilizadas para diferenciar texturas que não seguem um determinado padrão. Os Descritores de Haralick usam os seguintes descritores: homogeneidade, probabilidade máxima, entropia, momento de diferenças ordem k, momento inverso de diferença de ordem k, variância inversa, energia, contraste, variância, correlação, entre outros descritores.

Bihui e colaboradores (2010) elaboram um método de reconhecimento de madeira com base nos descritores de haralick extraídos da Matriz de Co-ocorrência de Níveis de Cinza com intervalo de pixels com 4 e nível de cinza com 128. Assim seis recursos: energia, entropia, contraste, dissimilaridade, momento da diferença inversa e variância, foram usados como recursos de classificação do experimento.

Hangjun (2012 - 2) apresentou um método de reconhecimento de madeira com base na entropia, um dos descritores de haralick, como parâmetro extraído do Filtro de Garbor. Os resultados experimentais mostraram melhorias na taxa de reconhecimento de madeira com a presença da entropia na tarefa de extrair características de textura com o auxílio do método Gabor Wavelet.

Por sua vez, no trabalho de Tou (2007), após a aquisição da imagem, aplicação do filtro passa alta¹⁰ e equalização de histograma utilizou-se os descritores de haralick. Desta forma calculou-se uma matriz de co-ocorrência utilizando os resultados como entrada de uma RNA. Para a realização dos experimentos foram utilizadas 360 imagens obtidas no Centro de

¹⁰ Filtro passa alta: tem como função estabelecer o valor zero em todas as frequências dentro de um círculo de raio definido enquanto que todas as frequências fora do círculo passam sem alteração.

Inteligência Artificial e Robótica. No primeiro experimento, obteve-se uma taxa de reconhecimento de 72%, e no segundo obteve-se o reconhecimento de 60%.

Outro método de utilização para segmentação de imagens de madeira é o PCA (do Inglês, *Principal Components Analysis* - PCA) um método clássico de redução de dimensão linear de dados, na análise do erro mínimo médio quadrático com menor dimensão de acordo com os dados originais, que se caracteriza por simplicidade e eficiência. O trabalho de Hangjun e equipe (2011) inova propondo algoritmos de identificação de faces, aplicados ao PCA para tratar o reconhecimento da imagem, amplamente utilizados na redução de dimensão em visão computacional.

A transformada de Fourier é uma ferramenta matemática que realiza a transição entre as variáveis de tempo e frequência de sinais, e decompõe um sinal em suas componentes elementares seno e cosseno (MITRA, 1999). Foi desenvolvido como aplicação inicial para resolução de problemas da condução do calor (lei da condução térmica), hoje amplamente utilizado no processamento de sinais e processamento de imagem. Método de segmentação como Quantização de Fase Local (LPQ) e Padrão Binário Local (LPB) baseiam-se em transformada de Fourier de quantização. Martins e equipe (2012) desenvolvem descritores estruturais de textura, utilizando Quantização de Fase Local (LPQ), mostrando por meio de experimentos que apresentam melhores resultados referentes ao Padrão Binário Local (LPB) e suas variantes. No entanto, segundo experimentos explorados na pesquisa, a combinação de ambos (LPB e LPQ) gera melhorias nos resultados, melhoria de cerca de 7 pontos percentuais, alcançando uma taxa de reconhecimento global de 86,47% em um banco de dados composto por 2240 imagens microscópicas, extraídas de 112 diferentes espécies florestais.

A possibilidade de utilizar vários métodos de segmentações para extrair o máximo possível de características e, conseqüentemente, melhorias na classificação comprovadas nas taxas de reconhecimento pode ser observada no trabalho de De Paula Filho (2012) integrando métodos como análises de cor, GLCM, histograma de borda, Fractais, LBP, LPQ e Gabor, com finalidade de extrair atributos da madeira.

De Paula Filho propõe o reconhecimento de espécies florestais através de imagens macroscópicas, e aplica duas formas de aquisição de imagens: abordagem tradicional em laboratório e abordagem em campo. Posteriormente estas imagens são divididas em sub-imagens a fim de que problemas locais não afetem a classificação geral da imagem. A partir delas, são extraídas informações de cor e de textura que são utilizadas para a construção de conjuntos de treinamento, teste e validação de classificadores. Com a abordagem em campo,

com presença de ruídos nos conjuntos de dados e sem tratamento, a taxa de reconhecimento obtida nessa etapa foi 95,82%, com a utilização da base de imagens criada com a abordagem tradicional em laboratório. Com esse novo modelo, a taxa de classificação foi de 99,49%.

3.1.3 Classificação

O reconhecimento de padrões é o campo da ciência que tem por objetivo a classificação de objetos em um determinado número de categorias ou classes a partir da observação de suas características (THEODORIDIS e KOUTROUMBAS, 2003).

Dado um conjunto de objetos com características mensuráveis, a classificação consiste em tentar categorizá-los. Tais características formam um espaço multidimensional (espaço de características), onde cada objeto é representado por um vetor de características (padrão), um ponto neste espaço. Assim, a tarefa da classificação pode ser geometricamente entendida como o reconhecimento de agrupamento no espaço de características.

A Tabela 6 apresenta os trabalhos correlatos com esta pesquisa, identificando seus respectivos métodos de classificação. Esta seção está ordenada pelos métodos de classificação mostrados na Tabela 6.

Tabela 6. Classificadores

Autor	KNN	RNA	SVM	Algoritmo de Otsu	Erro Médio Quadrático
(XUEBING, 2005)		X			
(BIHUI et al., 2010)			X		
(DE PAULA FILHO et al., 2014)			X		
(DE PAULA FILHO, 2012)		X			
(HANGJUN et al., 2009)				X	
(HANGJUN et al., 2011)	X		X		
(HENGNIAN et al., 2008)					X
(KHALID, 2008)		X			
(LINJIN et al., 2012)			X		
(MALLIK et al., 2011)	X	X	X		
(MARTINS et al., 2012)			X		
(PAULA e TUSSET, 2009)		X			
(LINGJUN et al., 2011-1)	X				
(LINGJUN et al., 2011-2)	X				

(TOU, 2007)		X			
(HUI, 2009)		X			
(SHAOCHUN, 2007)	X	X			
(HAIPENG, 2007)				X	
(ZHIWEI et al., 2011)			X		

Shaochun (2007) desenvolveu um método de extração de recursos da coloração da madeira, comparando pequenas diferenças nas cores com característica de isometria e alto poder de resolução, para classificação utilizou-se o método KNN aplicado a cinco espécies arbóreas.

Dentre os métodos mais aplicados, destacam-se as RNAs, baseado na sua abrangência em aplicabilidade e por dispor de classificação supervisionada e não supervisionada. Desta forma, Khalid (2008) apresenta um sistema de reconhecimento de espécies florestais tendo como base mais de 1900 imagens de 20 diferentes espécies presentes na Malásia. Foram utilizadas as redes neurais artificiais para classificação em grupamento por espécie. Para a aquisição das imagens foi utilizada uma câmera de padrão industrial e alto desempenho (JAI CV-M50) e para a iluminação foi desenvolvido um *led array*¹¹.

Por sua vez, Zhiwei et al. (2011) utilizam o método de classificação SVM como suporte ao desenvolvimento de um algoritmo de região de crescimento que atua como método de extração de características morfológicas dos poros da madeira. Este método de segmentação utiliza micrografias para adquirir dez características morfológicas das células dos poros da madeira. A experiência de simulação mostrou que este algoritmo poderia melhorar a velocidade computacional de segmentação de poros da madeira. As dez características morfológicas das células dos poros têm características bastante divergentes nos seis tipos de madeiras folhosas e, com isto, o algoritmo proposto demonstra alta eficiência no reconhecimento de espécies de folhas largas.

O método de Otsu é um algoritmo de limiarização, proposto por Nobuyuki Otsu (OTSU, 1975). Seu objetivo é, a partir de uma imagem em tons de cinza, determinar o valor ideal de um threshold que separe os elementos do fundo e da frente da imagem em dois grupamentos, utilizando a cor branca ou preta para cada um deles.

Hangjun e colaboradores (2009) propõem um método para identificar regiões de mata fechada, com segmentação através de variações entre classes, dividindo as regiões com o

¹¹ Led array: iluminação infravermelho com utilização para captura de imagem

algoritmo de Otsu, utilizando histograma de escala de cinza. Em outra aplicação, Haipeng (2007) utiliza o algoritmo de Otsu para reconhecimento botânico por características de textura de imagens de madeira, medindo parâmetros, como cor, saturação, iluminação, contraste, segundo momento angular, soma das variações, dimensão fractal e proporção de energia horizontal *wavelet*.

Por fim, como método de classificação, Hengnian e colaboradores (2008) propuseram um método de identificação de madeira baseado em uma análise de característica quantitativa dos poros, no qual aplica-se o Erro Médio Quadrático como forma de quantificar os poros da madeira para descrever a distribuição de poros e, desta forma, classificar a espécie botânica. Tal aplicação difere dos métodos tradicionais qualitativos, por meio de métodos de morfologia matemática, como a dilatação, erosão, transformação de seções transversais de madeira, imagem reparação, filtragem de ruído e detecção de borda para o segmento dos poros de seu fundo.

3.1.4 Taxas de Reconhecimento de Padrão

Os resultados da taxa de reconhecimento representam o equilíbrio entre a quantidade e a relevância das características abstraídas da imagem, integrado com a capacidade do método de classificação em reconhecer padrões com eficácia. Desta forma, a tabela 7 apresenta a taxa de reconhecimento de trabalhos relacionados a esta pesquisa.

Tabela 7. Taxa de reconhecimento

AUTOR	Taxa de reconhecimento %	Total imagens	Nº Espécies	Localidade
(XUEBING, 2005)	88	300	X	China
(BIHUI et al., 2010)	91,7	480	24	China
(DE PAULA FILHO et al., 2014)	97,7	2942	41	Brasil
(DE PAULA FILHO, 2012)	99,49	163	14	Brasil
(KHALID, 2008)	95	1949	20	Malásia
(MARTINS et al., 2012)	86,47	2240	112	Brasil
(PAULA e TUSSET, 2009)	80,9	163	14	Brasil
(LINGJUN et al., 2011-1)	93,3	X	X	China
(LINGJUN et al., 2011-2)	97,5	480	24	China
(TOU, 2007)	60 / 72	360	X	Cairo

No trabalho de Lingjun e equipe (2011-2) a taxa de reconhecimento alcançou limiares de 97,5% de eficiência. O método de reconhecimento de madeira foi desenvolvido com base em análise de textura onde as imagens da amostra foram divididas em várias regiões, após o corte de imagens estereograma madeira. Contudo, mais recursos foram extraídos pelo filtro de Gabor Wavelets¹² por cinco escalas e oito orientações. Para preservar regiões importantes da imagem, foram utilizados recursos de Gabor, clustering¹³ e operação de peneiramento para expor as características mascaradas em regiões de ruído, como a região de clivagem, a região do canal de resina e assim por diante.

Em De Paula Filho e colaboradores (2014) os índices da taxa de reconhecimento alcançaram limiares de 97,7% de eficiência. Este método identifica espécies botânicas, utilizando o classificador *SVMtrained* com um conjunto de recursos baseados em textura.

3.2 Ontologia Ambiental

Uma ontologia de domínio trata de um modelo conceitual que deve satisfazer o requisito adicional de servir como uma representação de consenso (ou modelo de referência) de conhecimento compartilhado por uma determinada comunidade. Portanto, se uma ontologia de domínio é, antes de qualquer coisa, um modelo conceitual, uma linguagem adequada para representação de ontologias de domínio deve satisfazer os requisitos gerais de uma linguagem adequada para modelagem conceitual.

Ontologias aplicadas ao cenário ambiental situado como domínio amplo e complexo, tornam-se mais representativas quando utilizadas como base para SEs de gestão, controle, monitoramento e fiscalização. Tal constatação baseia-se na observação dos numerosos e relevantes trabalhos relacionados a esta pesquisa, dispostos na Tabela 8. A descrição dos trabalhos correlatos segue a organização apresentada na tabela 8 que divide os trabalhos por área de aplicação, nomeadamente Manejo Florestal, Monitoramento e impacto ambiental, e Ecossistema e Biodiversidade.

Tabela 8. Aplicação da Ontologia

AUTOR	Manejo	Monitoramento e	Ecossistema e
-------	--------	-----------------	---------------

¹² Filtro de Gabor Wavelets: é um conjunto de funções Gaussianas moduladas por funções sinusoidais complexas, bidimensionais e apresentam certas propriedades extremamente úteis para fins de classificação de imagens (DAUGMAN, 1988).

¹³ Clustering: Agrupamento, dividir em grupos determinados valores, objetos.

	Florestal	Impacto Ambiental	Biodiversidade
(GU et al., 2004)	X		
(RENNOLLS, 2005)	X		
(SILVA, 2009)	X		
(QUINELATO; MORI e ROSI, 2008)	X		
(SOUZA, 2014)	X		
(MONTEIRO, 2006)		X	
(KAUPPINEN e ESPINDOLA, 2011)		X	
(BRANCO, 2013)		X	
(FORTES et al., 2008)		X	
(KALABOKIDIS, 2011)		X	
(SILVA, 2014)			X
(CAMPOS DOS SANTOS et al., 2011)			X
(MIYAZAKI, 2011)			X

3.2.1 Manejo Florestal

A extensa diversidade biológica da Amazônia e a vasta quantidade de espécies florestais existentes dificulta a árdua tarefa de identificação das plantas pelos técnicos que atuam no manejo florestal. Outra dificuldade é abundância de nomes vernaculares e a variação destes nomes de região para região.

A solução para os problemas que ocorrem no manejo florestal é incorporar ao projeto de manejo florestal um caráter científico. Neste sentido, trabalhos como o de Rennolls (2005) propõem uma ontologia no campo do plano de manejo florestal, mais especificamente na mensuração e inventário florestal, que visa integrar sistemas que atuam na área a fim de criar um padrão de modelo florestal.

Outro trabalho, proposto por Gu e colaboradores (2004), desenvolve uma ontologia de domínio específico para área de botânica, a qual contém uma estrutura ontológica multiperspectiva da botânica, a fim de organizar os conceitos botânicos e conhecer as

diferentes necessidades dos usuários. Esta ontologia apresenta um conjunto de axiomas bem definidos, que são usados para verificação e raciocínio a partir do conhecimento adquirido, analisados neste trabalho pelos quesitos consistência, integridade e redundância.

Qualquer violação de axiomas ontológicos é um pedaço de evidência de inconsistência na base de conhecimento. A análise de redundância para a base botânica conhecimento é também baseada em ontologias. Não há necessidade de manter os pedaços de conhecimento que podem ser derivados de outros conhecimentos e axiomas ontológicos (GU, 2004).

3.2.2 Monitoramento e Impacto Ambiental

Segundo o Artigo 1º da Resolução n.º 001/86 do Conselho Nacional do Meio Ambiente (CONAMA), Impacto Ambiental é "qualquer alteração das propriedades físicas, químicas, biológicas do meio ambiente, causada por qualquer forma de matéria ou energia resultante das atividades humanas que afetem diretamente ou indiretamente: a saúde, a segurança, e o bem estar da população; as atividades sociais e econômicas; a biota; as condições estéticas e sanitárias ambientais; e a qualidade dos recursos ambientais" (CONAMA, 1986).

Por sua vez, o monitoramento ambiental permite ainda, compreender melhor a relação entre ações do homem e o meio ambiente, bem como o resultado da atuação das instituições por meio de planos, programas, projetos, instrumentos legais e financeiros, capazes de manter as condições ideais dos recursos naturais (equilíbrio ecológico) ou recuperar áreas e sistemas específicos (RAMOS e LUCHIARI, 2016).

Neste sentido, são apresentados a seguir alguns trabalhos científicos aplicados ao âmbito do monitoramento e impacto ambiental que integram GC para fins de melhorias do cenário ambiental, e consequentemente estão relacionados a esta pesquisa.

Monteiro (2006) desenvolve a Ontologia Avaliação do Ciclo de Vida (ACV) para fomentar a disseminação dos conceitos do domínio ACV, propondo assim uma ferramenta de gestão ambiental.

No trabalho de Kauppinen e Espindola (2011) foi proposta a ontologia *Pluto* para integrar semanticamente e orientar o processo de raciocínio sobre os conjuntos de dados relacionados ao desmatamento e a trajetória de mudança de uso de terra na Amazônia brasileira.

Em Branco (2013), propõe-se uma ontologia para suporte ao levantamento e à avaliação de ocorrências de impactos ambientais, integrado ao sistema de gestão ambiental

(conforme a NBR ISO 14001:2004) aplicado ao desenvolvimento urbano, mais especificamente ao empreendimento de engenharia.

3.2.3 Ecossistema e Biodiversidade

Dados científicos, de maneira geral, encontram-se geralmente dispersos em diferentes fontes, algumas de fácil acesso (via Web), periódicos e livros científicos, relatórios técnico-científicos, dissertações e teses, e outras de difícil localização e acesso, como arquivos, pastas e cadernos de campo. Algumas dificuldades devem-se também à falta de uma política para gestão destes dados e do potencial conhecimento por eles propiciado. É visível a crescente demanda por estes dados em diversas aplicações consideradas importantes, como avaliação de impacto ambiental, definição de áreas de preservação ambiental, proteção de espécies ameaçadas, recuperação de áreas degradadas, bioprospecção, estabelecimento de políticas públicas, legislação ambiental, entre outras (CAMPOS DOS SANTOS et.al, 2011).

Os dados e o conhecimento científico sobre a biodiversidade exercem um importante papel no atendimento às demandas deste tipo, pois acumulam investimentos de anos em exploração e pesquisa. No entanto, tornar isso cada vez mais acessível ao público de forma adequada, rápida e confiável, impõe o desenvolvimento de sistemas de informação capazes de extrair, armazenar, gerenciar, analisar, integrar e disseminar os diferentes dados das diversas fontes de dados de biodiversidade (CAMPOS DOS SANTOS, BY e MAGALHÃES, 2000)(UMMINGER e YOUNG, 1997).

Neste contexto, os trabalhos científicos situados no âmbito de Ecossistemas e Biodiversidade correlatos a esta pesquisa aplicam GC, utilizando ontologia, para estruturar o conhecimento angariado por especialistas, como se pode observar em:

- Silva (2014): desenvolve uma ontologia funcional FOBiOS para prover a interoperabilidade entre os padrões ABCD e Darwin Core¹⁴. A ontologia desenvolvida está disponível para uso por serviços de tradução e mapeamento dos dados descritos nesses dois padrões.
- Santos e colaboradores (2011): baseado em estudos de casos do INPA, propõem aplicar tecnologias da Web Semântica, como metadados, modelos e ontologias, aos problemas de interoperabilidades de dados e sistemas sobre a biosfera e atmosfera da Amazônia.

¹⁴ Padrões ABCD e Darwin Core: os padrões de dados definidos por estes grupos abrangem esquemas de dados e protocolos de transferência de dados. São dois os padrões mais utilizados para coleções de dados primários: Access to Biological Collections Data (ABCD) e Darwin Core (DwC) (SILVA, 2014).

- Miyazaki (2011): propõe uma ontologia de coleções biológicas, fornecendo uma única estrutura unificada para o conhecimento da biodiversidade, com intuito de melhorar a agregação e integração de dados em todo o domínio da biodiversidade, que se estendem por genética do organismo, e os níveis de organização dos ecossistemas. O autor argumenta que, se adotada como um padrão e rigorosamente aplicada e enriquecida pela comunidade biodiversidade, esta ontologia reduziria significativamente as barreiras à descoberta de dados, integração e intercâmbio entre os recursos da biodiversidade e pesquisadores.

3.2.4 Tecnologias Associadas

Elementos tecnológicos permeiam todas as atividades científicas e com esta área não é diferente. Várias tecnologias são encontradas com relativa frequência nos trabalhos correlatos e são elas que figuram na Tabela 9. É visível também os casos onde várias tecnologias estão presentes no mesmo trabalho, o que sugere alguma forma de integração ou colaboração entre elas.

Tabela 9. Algumas tecnologias usadas

AUTOR	Ontologia	SIG	Mineração de dados	Sistemas Multiagentes
(GU et al., 2004)	X			
(RENNOLLS, 2005)	X			
(SILVA, 2009)	X		X	
(QUINELATO; MORI e ROSI, 2008)	X	X		
(SOUZA, 2014)	X	X		
(MONTEIRO, 2006)	X			
(KAUPPINEN e ESPINDOLA, 2011)	X			
(BRANCO, 2013)	X			
(FORTES et al., 2008)	X			X
(KALABOKIDIS, 2011)	X	X		
(SILVA, 2014)	X			
(SANTOS et al., 2011)	X			
(MIYAZAKI, 2011)	X			

Silva (2009) integra ontologia à mineração de dados no desenvolvimento do sistema de gestão e monitoramento da biodiversidade mais especificamente a GC taxonômico aplicado na conservação da flora brasileira, no intuito de facilitar o manuseio dos respectivos dados taxonômicos, tendo em vista que novas espécies de plantas são descobertas, necessitando de identificação e as constantes atualizações dos nomes por especialistas. Tal sistema visa definir uma metodologia para gerenciar e partilhar o conhecimento destes pesquisadores. Desta forma foi realizado um estudo do uso das tecnologias de colaboração, ontologias e mineração de dados possibilitando a GC na conservação da flora brasileira.

O trabalho de Quinelato, Mori e Rosi (2008) propõe o OntoSIGF, uma integração entre SIG e ontologia aplicada à área florestal, onde a ontologia tem como papel registrar e organizar o conhecimento do domínio geográfico de uma empresa que trabalha com manejo de áreas florestais, através da descrição de seus conceitos, relações e demais características, elaborando um processo de busca e geração de mapas da cartografia baseado na ontologia proposta.

Souza (2014) apresenta a integração de ontologias, formando assim uma rede ontológica que servirá como base para a publicação e ligação de dados sobre madeira, comunidades e plantas do bioma cerrado brasileiro, obtidos a partir de estudos científicos, dados meteorológicos, ambientais e informação geográfica (mapas). Foram utilizadas as bases de conhecimento “Dados Cube”, “Meteorologia”, “GeoSPARQL” como partes integrantes da infraestrutura ontológica.

A integração da ontologia com os Sistemas Multiagentes no trabalho de Fortes e colaboradores (2008) apresenta um suporte para o monitoramento do complexo portuário da ilha de São Luís (MA). O estudo trata do suporte de agentes inteligentes no processo de análise bioquímica da atividade das enzimas de destoxificação e de outras enzimas de ação antioxidante. No término do estudo se evidenciaram anomalias no comprometimento natatório de algumas espécies de peixes da região portuária proveniente de ingestão de toxinas, acarretando um desequilíbrio aquático.

Por sua vez, Kalabokidis (2011) desenvolve o sistema OntoFire, um geo-portal sobre incêndios florestais, baseado em ontologia. A abordagem proposta visa melhorar a descoberta de informações valiosas, que são necessárias para definir prioridades e estratégias de mitigação de desastres e prevenção.

3.3 Sistemas especialistas e chaveamentos no processo de identificação botânica

O processo de identificação botânica com auxílio de ferramentas computacionais apresenta um longo histórico de aplicações. Os primeiros trabalhos apresentam métodos rudimentares de classificação, utilizando chaveamento como pode-se observar nos trabalhos de Morse e equipe (1968) aplicado à flora Norte Americana. Contudo trabalhos mais recentes têm apresentado métodos de chaveamento, como as técnicas de identificação botânica observado no trabalho de Branch e colaboradores (2005) que utiliza a base interativa da web como instrumento para um software de identificação de uma vasta quantidade de espécies de localidades diversas.

O desenvolvimento de SEs no cenário ambiental mais especificamente aplicado ao processo de identificação botânica avançam para novos patamares de pesquisa como pode-se observar no trabalho de Woolley e equipe (1987) com a construção de um sistema especialista SYSTEX – um sistema inteligente baseado em regras com tolerância à falta de dados e à falhas. Posteriormente, outros trabalhos na área das ciências ambientais exploram este novo paradigma da inteligência artificial, como por exemplo o trabalho de Bailón e colaboradores (1993) que desenvolveu o sistema especialista GREEN (*Gymnosperms Remote Expert Executed Over Networks*) no auxílio à tomada de decisão no processo de identificação botânica.

O sistema especialista proposto neste trabalho contempla um domínio não explorado, tanto nos aspectos computacionais, posto que a base de conhecimento e a máquina de inferência estão conectadas a ontologia ONTO-AmazonTimber aplicada ao contexto botânico, como no campo de abrangência do cenário aplicado, uma vez que o cenário botânico restringe-se à identificação botânica de espécies florestais amazônicas comercializadas no setor madeireiro.

3.4 Análise Comparativa

A busca pelo aumento da acurácia da identificação botânica utilizando recursos tecnológicos é cada vez mais parte integrante do plano de manejo florestal ou da fiscalização no cenário ambiental da indústria madeireira, ao mesmo tempo em que os avanços na pesquisa são evidenciados pelo volume de publicações apresentados no referencial teórico. Contudo ainda existe a necessidade de melhorias em nível tecnológico, aplicações que possam

abranger as espécies da Amazônia, levando em consideração o contexto problemático do contrassenso dos nomes vernaculares com os nomes científicos.

Neste contexto foca-se um dos temas deste trabalho, que é suportar a construção do conhecimento necessário para desenvolver um referencial semântico no domínio da botânica, capaz de identificar espécies botânicas colhidas no manejo florestal sustentável na Floresta Nacional do Tapajós.

Dentre as contribuições apresentadas neste trabalho, destaca-se a integração da GC (através de um referencial semântico) com técnicas modernas de identificação. Tal integração trará as vantagens inerentes à GC, possibilitando um recurso de identificação botânica que leva em consideração os conhecimentos dos agentes envolvidos como mateiros e taxonomistas. Inova ainda com os cenários de aplicação:

- A ontologia proposta abrange um domínio ainda não explorado na literatura, que aborda a prática específica de identificação botânica aplicada às espécies da Amazônia. Observa-se na literatura uma abordagem geral como apresentado nos trabalhos de (SILVA, 2009) abrangendo toda flora brasileira como no trabalho de (GU et al., 2004) de forma genérica aborda sobre o domínio da botânica, e a aplicação nas mais diversas áreas como incêndios em áreas florestais (KALABOKIDIS, 2011), botânica com auxílio de Sistemas de Informações Geográficas (QUINELATO, 2008), entre outras aplicações apresentadas no referencial teórico. O trabalho de Rennols (RENNOLLS, 2005) apresenta uma similaridade por se tratar de uma ontologia aplicada ao manejo florestal no qual inclui inventário florestal, mas não trata do processo de identificação botânica;
- Identificação botânica por imagem da anatomia da madeira: utiliza-se como características da segmentação a texturização observados nos trabalhos de (TOU, 2007) (SANTOS, 2009) (CASANOVA, 2008) (LINJIN et al., 2012) (HANGJUN et al., 2012). Na segmentação da imagem, serão extraídos aspectos da matriz de co-ocorrência de níveis de cinza utilizando como parâmetro as características de *haralick* observadas nos trabalhos de (BIHUI et al., 2010) (TOU, 2007); e
- Identificação botânica com um suporte à tomada de decisão utilizando uma base de conhecimento ontológico: observa-se que vários dos referenciais semânticos existentes na literatura não apresentam uma aplicação específica para o cenário amazônico, menos ainda quando se trata de identificação botânica de espécies exploradas no manejo florestal.

4. METODOLOGIA DA PESQUISA

Este capítulo destina-se a apresentar as delimitações metodológicas da pesquisa, isto inclui as metodologias aplicadas, os métodos e procedimentos utilizados para o desenvolvimento de um referencial semântico no âmbito da identificação botânica de espécies amazônicas.

4.1 Método clássico

Segundo Lakatos e Marconi (1983), uma pesquisa é um processo cognitivo que contém formalismo, regras e padrões a serem seguidos, utiliza-se do método científico¹⁵ para condução e delineamento do percurso da pesquisa.

A pesquisa aqui apresentada fundamenta-se no empirismo indutivista¹⁶ de Francis Bacon (1561 – 1626), o qual aplica o método científico tradicional: partindo de observações sistemáticas, formulação de hipóteses, seguido de experimentações e conclusões.

A prática da experimentação seguida pela corrente empirista vem sendo aprimorada no decorrer da história da ciência. Desta forma, a Figura 21 ilustra uma visão mais cotidiana do método científico tradicional.

¹⁵ Para Gil (2006), o método científico trata-se de um arcabouço de procedimentos, técnicas e atividades sistemáticas e racionais para atingir de forma efetiva o objetivo da pesquisa: a produção do conhecimento, ampliando reflexos e pressupostos para novas pesquisas.

¹⁶ A abordagem empírico-indutiva tem como preceitos a concepção do conhecimento por de experiências da realidade material que originam-se de fora para dentro. O pesquisador atua como expectador dos fenômenos, mas seguindo preceitos como objetividade, neutralidade na observação (GALLIANO, 1979)

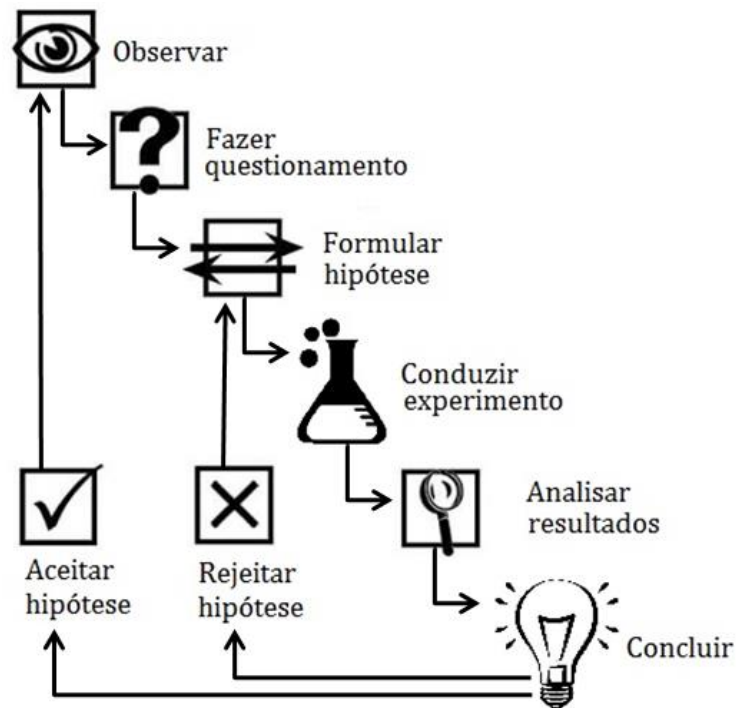


Figura 21. Etapas do método científico

O método científico tradicional inicia-se com a observação de um fenômeno que mobiliza o interesse investigativo e, por conseguinte, questionamentos surgem acompanhados por afirmativas de cunho dedutivo representado por hipóteses. Experimentos são conduzidos para testar repetidamente em diversas condições as hipóteses formuladas. Após a análise dos resultados, pode-se concluir sobre a aceitação da hipótese ou rejeição para posterior aprimoramento ou reformulação.

Nesta tese, observa-se a instanciación do método científico clássico no desenvolvimento global da tese, o que está refletido nos seguintes capítulos deste documento:

- Observação, questionamentos, hipóteses apresentados no capítulo introdutório;
- Experimentos e análise dos resultados mostrados nos capítulos 4 (Metodologia da Pesquisa) e 5 (Referencial semântico aplicado ao processo de identificação botânica)
- Conclusões expostas no capítulo 6.

4.2 Metodologias Complementares

4.2.1 Metodologia para o desenvolvimento da Ontologia

Metodologias de desenvolvimento de ontologias existem no intuito de sistematizar a formalização do conhecimento de especialistas, técnicos ou equipes de projetos. Segundo

Lima e colaboradores (2005), as metodologias divergem quanto ao foco de aplicação, algumas destas metodologias têm como alvo a captação de feedback; outras metodologias têm como foco a modelagem empresarial; existem metodologias que visam melhorar a expressão de requisitos tácitos ou implícitos. No entanto as mais observadas na literatura atuam na estrita construção e manutenção de projetos de gestão do conhecimento, a citar: Methontology (FERNANDEZ, GOMEZ-PEREZ e JURISTO, 1997); On-to-Knowledge (STAAB, SCHNURR, STUDER e SURE, 2001); Enterprise (JONES et al., 1998) metodologia de Gruninger e Fox (GRUNINGER e FOX, 1995); método de Uschold e King (USCHOLD e KING, 1995); método Cyc (REED e LENAT, 2002); método Kactus (BERNARAS, LARESGOITI e CORERA, 1996); método Sensus (SWARTOUT et al., 1997); e g) método 101 (NOY e MCGUINNESS, 2001).

Contudo, nenhuma das metodologias descritas apresenta total estabilidade, abrangência, principalmente se comparadas com metodologias de engenharia de software (ALMEIDA, 2003).

A metodologia aplicada neste trabalho advém do projeto e-COGNOS (Lima, El-Diraby e Stephens, 2005) no qual foi desenvolvida uma plataforma de gestão do conhecimento baseada na web, que tem como alvo inicial as necessidades da indústria do sector da construção, onde foi desenvolvido a plataforma computacional e-CKMI (*e-COGNOS Knowledge Management Infrastructure*) para representação do conhecimento e construção de itens do conhecimento.

Mais especificamente a ontologia e-COGNOS é composta por duas taxonomias, nomeadamente uma taxonomia de conceitos e uma taxonomia de relações. Apresenta uma representação semântica na qual “Conceitos” e “Propriedades” são especializações da classe *Object*; para caracterizar um conceito, o meta-modelo e-COGNOS define a classe *Attribute*, que é uma especialização da classe *Relation* (LIMA et al., 2005).

O método proposto para o desenvolvimento da ontologia foi inspirado na abordagem usada pelo projeto e-COGNOS. Assim como os principais conceitos que servem como a espinha dorsal da ontologia também foram inspirados na ontologia e-COGNOS. No entanto, para o propósito específico desta tese, algumas adaptações e refinamentos do modelo ontológico tiveram de ser feitas, em conformidade com o trabalho de Ruben (2014). O método aqui adotado usa uma abordagem iterativa (Figura 22), que é dividida em várias fases, com cada fase contendo um conjunto de tarefas relacionadas:



Figura 22. Metodologia e-COGNOS para construção da Ontologia

Fonte: Adaptado de Costa (2014)

4.2.1.1 METODOLOGIA PARA VALIDAÇÃO DO REFERENCIAL SEMÂNTICO – ONTOLOGIA

A metodologia para validação considera um processo que fornece evidências que avaliem se o referencial semântico está adequado ao propósito para o qual foi concebido. Neste sentido, a metodologia para validação do referencial semântico aqui aplicada objetiva avaliar os resultados decorrente das inferências e das inserções sobre o referencial semântico, por meio de sistemas computacionais. Para tal, desenvolveram-se dois sistemas especialistas no intuito de reconhecer padrões do referencial semântico: um sistema especialista para identificação de espécies amazônicas baseados no reconhecimento de padrões de características da espécie e do contexto do domínio geográfico em que está inserida; e um sistema especialista para identificação de espécies amazônicas baseado no reconhecimento de padrões em amostras de imagens de madeira.

4.2.1.2 LINGUAGENS PARA ONTOLOGIA

Segundo Horridgen e equipe (2004), as linguagens para o desenvolvimento da ontologia divergem quanto aos recursos e às facilidades de programação. Entretanto, o desenvolvimento mais recente em padrão para linguagens aplicadas a ontologia é a *Ontology Web Language (OWL)* da W3C, considerada padrão na formalização de ontologias, que incorpora XML (do inglês, *eXtensible Markup Language*), RDF (do inglês, *Resource Description Framework*) e *RDF-Schemas*.

Segundo Horrocks e colaboradores (2003), a sintaxe da linguagem OWL tem como base o XML e apresenta vantagens quanto à disponibilidade de operadores lógicos, tags,

anotações. É possível representar propriedades transitivas, simétricas, funcionais e inversas, assim como é possível relacionar instâncias e conceitos, e as diferenças entre os indivíduos.

Neste contexto, em função dos recursos que a linguagem OWL apresenta, aliado à compatibilidade proferida entre as ferramentas e a metodologia para construção da ontologia, optou-se por utilizá-la como linguagem para o desenvolvimento da ontologia deste trabalho.

4.2.1.3 FERRAMENTA COMPUTACIONAL PARA DESENVOLVIMENTO DA ONTOLOGIA

Neste trabalho, utiliza-se a ferramenta *Protégé* para construção da ontologia, o *editor*¹⁷ de ontologias de popularidade entre as comunidades envolvidas com desenvolvimento de ontologias e que foi desenvolvido pelo *Knowledge Modeling Group* (KMG) da *Stanford Medical Informatics* na escola de medicina da universidade de Stanford (Califórnia, EUA). Esta ferramenta dispõe de uma interface gráfica intuitiva que permite a construção de ontologias de domínio, tanto por desenvolvedores de sistema como por especialistas em domínio (Figura 23) (PROTÉGÉ PROJECT, 2010).

A ferramenta Protégé tem seu código aberto, contém uma vasta quantidade de plugins, além de possível importação e exportação de códigos fontes de diversas linguagens como: Flogic, Jess, OIL, SML e Prolog. Entre outras vantagens, a ferramenta permite acesso à bases de dados, inclusão de imagens, anotações, criação e execução de restrições e fusão de ontologias (NOY, FERGERSON e MUSEN, 2000).

¹⁷ Editor: ferramenta computacional contendo interface gráfica com finalidade de suporte a atividades fins.

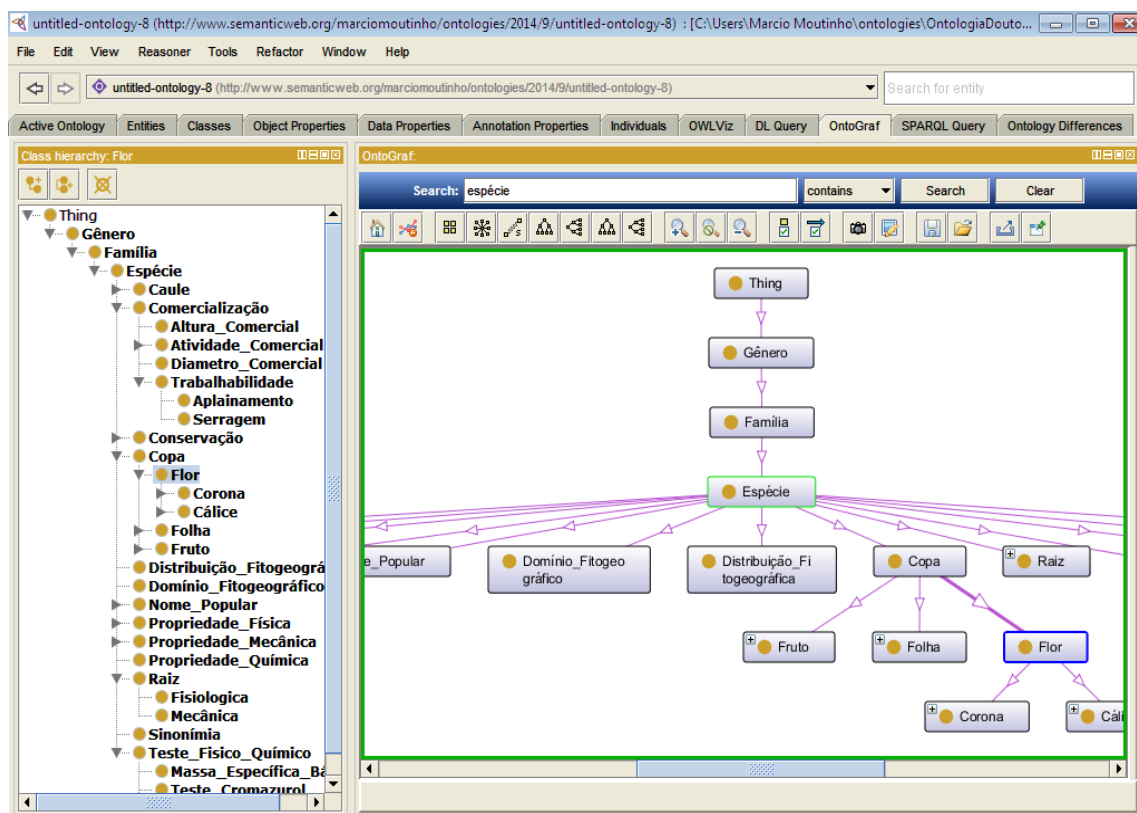


Figura 23. Ferramenta Protégé

4.2.2 Metodologia para o desenvolvimento dos Sistemas Especialistas

Os sistemas especialistas desenvolvidos neste trabalho foram construídos com uma abordagem de desenvolvimento iterativo, representado na Figura 24 (LA SALLE et al., 1990).

A instanciación desta metodologia é apresentada no capítulo 5 com o desenvolvimento do sistema especialista para classificação de imagem de madeira e do sistema especialista para identificação botânica – Inventário florestal.

Segundo Millette (2012) as partes interessadas durante as etapas de desenvolvimento de um sistema especialista são: a equipe de desenvolvimento, o engenheiro do conhecimento e os especialistas do domínio do conhecimento (Figura 24). A equipe de desenvolvimento visa codificar os componentes do sistema especialista projetados pelo engenheiro do conhecimento. O engenheiro do conhecimento atua como um mediador entre os especialistas do domínio de problemas e a equipe de desenvolvimento, e tem como função abstrair o conhecimento dos especialistas para compor a estrutura da base de conhecimento. O especialista do domínio do conhecimento tem como função dispor seu conhecimento por meio de execução de atividade, reportar experiência, fatos, habilidades e práticas de soluções sobre

o domínio de conhecimento. O usuário final do sistema computacional não necessariamente precisa ser um especialista, no entanto, caso seja um especialista, aumenta o nível de profundidade na caracterização das espécies, o que repercute nos resultados do sistema computacional, no entanto de forma geral o sistema atua como auxiliar na execução da tarefa e no diagnóstico da identificação da espécie.

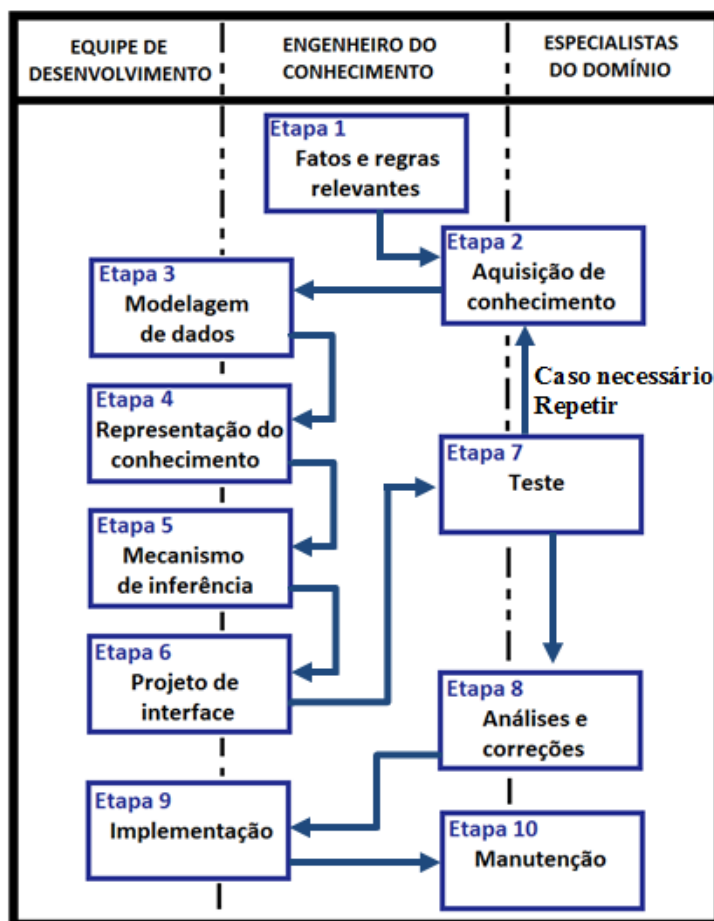


Figura 24. Metodologia do desenvolvimento dos sistemas especialistas.

Adaptado (LA SALLE et al., 1990)

O método inicia-se como a aquisição de literatura, fatos e regras relevantes (etapa 1). O engenheiro de conhecimento deve começar a recolher informações e entender a terminologia específica do domínio antes de realizar quaisquer reuniões formais ou entrevistas com os especialistas do domínio do conhecimento, tornando-se mais aclimatado com a terminologia específica do domínio, o que permite uma melhor comunicação entre o engenheiro de conhecimento e os especialistas do domínio do conhecimento.

A aquisição de conhecimento (etapa 2) situa-se como a etapa mais importante no processo de desenvolvimento do sistema especialista. O objetivo, durante esta etapa, é reunir

os requisitos de sistema para documentar o conhecimento necessário e assim construir a base de conhecimento. O engenheiro do conhecimento deve reunir de forma abrangente, completa e com profundidade o conhecimento dos especialistas do domínio da botânica que será necessário para identificar as espécies florestais. A integralidade e a exatidão da base de conhecimento é um fator crítico para o sucesso de um conhecimento baseado no sistema especialista.

O conhecimento científico necessário para sua construção provém da extração das seguintes fontes: literatura científica, entrevista com taxonomistas. Por sua vez, o conhecimento empírico provém de habilidade e experiências, capturadas por meio de atividades práticas com percepção de características experimentais de mateiros e taxonomistas.

Após a aquisição, inicia-se a modelagem dos dados (etapa 3) na qual aplicam-se técnicas de modelagem de dados, para construir uma estrutura de dados necessária para abrigar a informação permanentemente. Esta etapa serve para sistematizar o conhecimento angariado para posterior etapa de representação do conhecimento.

A representação do conhecimento (etapa 4) é definida como processo de estruturação do conhecimento para que possa ser acessado por um sistema computacional (HOPLIN e ERDMAN, 1990). Portanto, a aquisição de conhecimento, modelagem de dados e fases de representação do conhecimento devem ser realizadas em conjunto e de forma iterativa até que a base de conhecimento represente a amplitude completa e a profundidade de conhecimento no domínio do problema.

O conhecimento devidamente formalizado deve ser submetido à aplicação de um mecanismo de inferência (etapa 5) capaz de inferir regras sobre base de conhecimento a fim de produzir uma solução. O mecanismo de inferência deve funcionar seguindo as exigências e normativas dos especialistas do domínio do conhecimento angariadas pelo engenheiro do conhecimento. Contudo, para o bom funcionamento do motor de inferência, algumas características devem ser observadas, como: integridade, acurácia e tempo de resposta (GROSAN e ABRAHAM, 2011).

O conhecimento inferido deve ser analisado pelo usuário. Desta forma, o sistema especialista deve apresentar um projeto de interface (etapa 6) que possibilite o usuário a imergir sobre o domínio da botânica de forma intuitiva, permitindo apresentar um problema, obter soluções que o auxiliem no processo de tomada de decisão no processo de identificação botânica.

As etapas posteriores são chamadas de suporte: Teste (etapa 7), Análise e correções (etapa 8), Implementação (etapa 9) e Manutenção (etapa 10), tem como função manter a acurácia e bom funcionamento do sistema especialista.

A etapa de teste visa comprovar a exatidão do sistema especialista, com auxílio de técnicas que validem as ações dos sistemas especialistas. Após os testes, os resultados devem ser analisados e, em caso de discrepância com as práticas dos especialistas, novas incursões devem ser aplicadas ao projeto, reiniciando na etapa 2 de aquisição de conhecimento, mas com direcionamentos específicos para as discordâncias identificadas.

Na etapa de implementação, a equipe de desenvolvimento codifica os componentes do sistema especialista, integra a base de conhecimento ao mecanismo de inferência e por sua vez integra com a interface do usuário.

Os domínios de conhecimento onde um sistema especialista atua raramente são estáticos. Desta forma, os sistemas especialistas devem acompanhar a evolução do domínio de conhecimento. Caso mudanças do domínio do conhecimento não sejam mapeadas e incorporadas na base de conhecimento, esta se tornará estagnada, ultrapassada, ineficaz e possivelmente incorreta. Neste sentido, a manutenção deve ocorrer em duas diretivas: na primeira inicia-se um retorno à etapa 2 de aquisição de conhecimento o que torna necessária a intervenção do engenheiro de conhecimento; na segunda diretiva, os especialistas do domínio devem analisar constantemente novas condições introduzidas no domínio e ajustar a base de conhecimento em conformidade.

4.3 Procedimentos Metodológicos

4.3.1 Ambiente de Pesquisa: Amazônia - Tapajós

A pesquisa ocorre na Floresta Nacional do Tapajós, uma unidade de conservação criada pelo Decreto nº 73.684 de fevereiro de 1974, com uma área aproximada de 545 mil hectares, administrada pelo Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio). Localizada no oeste do Pará, nos municípios de Belterra, Aveiro, Rurópolis e Placas, Faz limite com o Rio Tapajós, com a rodovia BR 163-Santarém-Cuiabá e com o rio Cupari.

Mais especificamente, a identificação botânica ocorre no quilômetro 127 da rodovia BR 163, Santarém – Cuiabá, visto que apresenta uma base de pesquisa e uma identificação prévia em nível de inventário florestal, propiciando melhor investigação dos profissionais envolvidos.

A Figura 25 ilustra a abrangência da Floresta Nacional do Tapajós, apresentando paisagens ao longo do rio Tapajós bastante diversificadas, com igapós, capoeiras, seringais e florestas altas. A Floresta Nacional do Tapajós contém diversas espécies de valor comercial, como o cedro, a copaíba, a seringueira, castanheira, entre outras distribuídas sobre relevos planos e ondulações suaves, que se acentuam nos interflúvios da margem direita do Tapajós, ao sul da Unidade de Conservação (ICMBio, 2017).

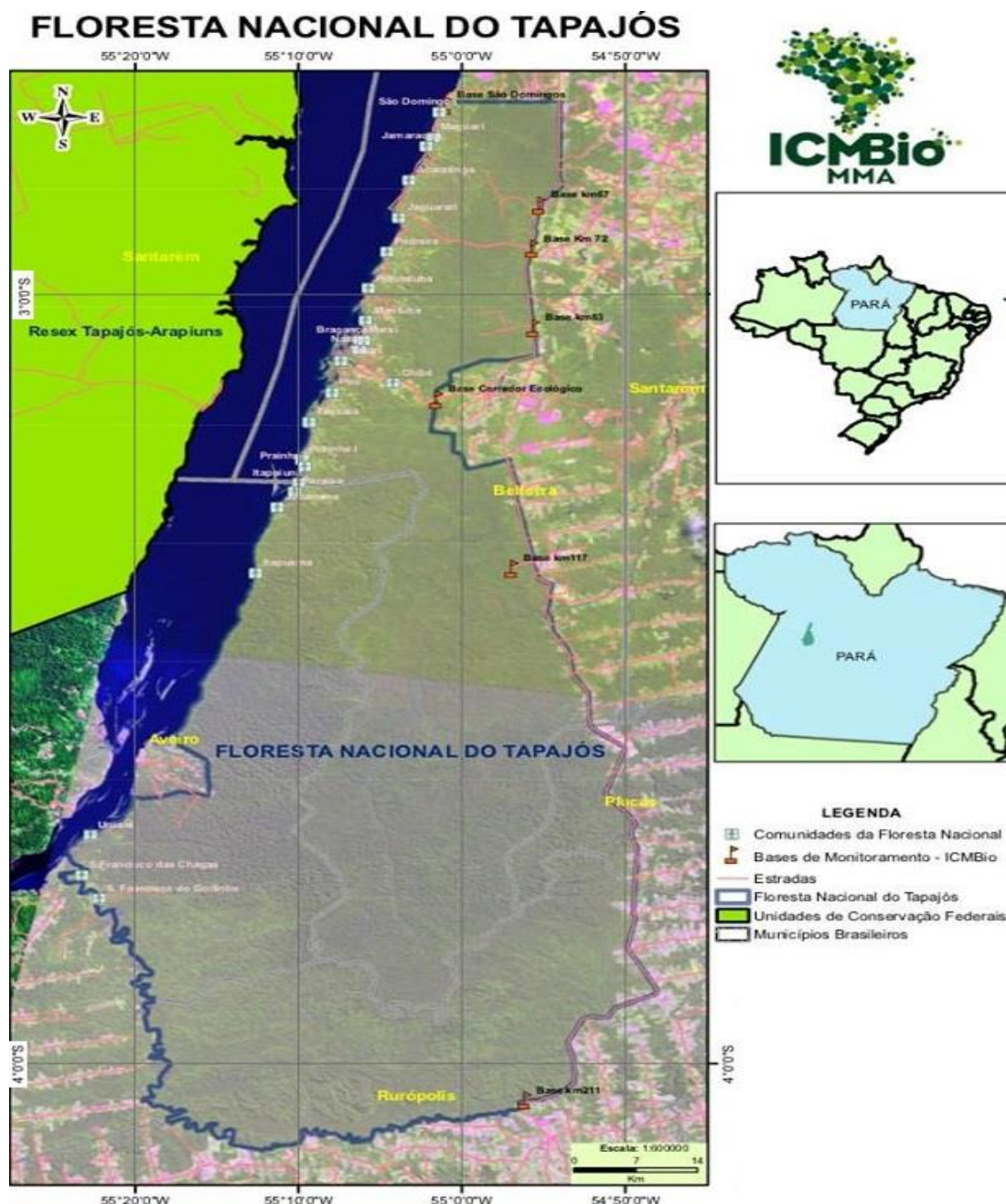


Figura 25. FLONA Tapajós, adaptado (ICMBio, 2017)

4.3.2 Levantamento das principais espécies comercializadas na região amazônica

Foi realizado, perante o IBAMA, SFB (Serviço Florestal Brasileiro), SEMA (Secretaria de Meio Ambiente do Pará) e demais órgãos competentes, o levantamento das principais espécies comercializadas na região do Oeste do Pará. Em seguida, por meio de literaturas especializadas e acesso ao banco de dados do Herbário da Embrapa Amazônia Oriental e Laboratório de Produtos Florestais – LPF do IBAMA, verificou-se a existência de dados referentes às características botânicas destas espécies.

Obtida a lista das espécies mais comercializadas, foram selecionadas as principais espécies (maior volumetria) para serem utilizadas como estudo de caso. Foram ainda incluídas algumas espécies que estão em extinção ou proibidas por lei sua comercialização a fim de aumentar as possibilidades de identificação.

4.3.3 Caracterização das amostras de madeira

Quanto à seleção de espécies, realizou-se levantamento da volumetria das madeiras comercializadas no estado do Pará entre 2007 – 2015 (SISFLORA/PA, 2016). Posteriormente, com o auxílio de especialista em identificação de madeiras amazônicas, associado com a disponibilidade *in loco* de material, vinte amostras de cada espécie selecionada, a citar *Bagassa guianensis* (tatajuba), *Carapa guianensis* (andiroba), *Cedrelinga cateniformis* (cedrorana), *Dipteryx ferrea* (cumarú), *Goupia glabra* (cupiúba), *Handroanthus* sp. (Ipê), *Hymenolobium petraeum* (angelim), *Manilkara* spp. (massaranduba), *Peltogeny* spp. (Pau-roxo), *Vataireopsis* sp. (fava) foram coletadas em diferentes serrarias de Rurópolis, Estado do Pará – Brasil. As amostras foram desdobradas em corpos de prova de 2,5 x 2,5 x 5 cm de espessura, largura e comprimento, respectivamente. Ressalta-se que as amostras foram, posteriormente, devidamente identificadas por especialistas em xiloteca.

Os corpos de prova foram polidos por meio de uma lixadeira de cinta portátil (Figura 25), nas granulometrias de 40, 80 e 120 mesh. Em cada fase do polimento foram obtidas imagens utilizando microscópio digital (Figura 26) com acesso via USB com câmera de 2.0 megapixel, modelo *Eletronic Magnifier*, com ampliação padrão de 500x. As imagens apresentam 600 x 600 pixels no espectro visível em um quadrante de amostra, medindo 0,5 x 0,5 cm como elemento de escala para a caracterização de dimensões da madeira. Para tal, a distância entre a amostra de madeira e o microscópio digital na captura da imagem mede

aproximadamente 7 cm. Assim, trabalhou-se com um total de 600 imagens, sendo 60 por espécie e 20 por tratamento.



Figura 26. Lixadeira de cinta portátil



Figura 27. Microscópio digital



5. REFERENCIAL SEMÂNTICO APLICADO AO PROCESSO DE IDENTIFICAÇÃO BOTÂNICA

Este capítulo apresenta a descrição das etapas necessárias para o desenvolvimento do referencial semântico iniciadas com o modelo conceitual, perpassando por etapas de projeto e implementação, seguindo para a validação e a análise dos resultados.

5.1 Definições e conceitos utilizados neste trabalho

Embora as definições e os conceitos apresentados neste tópico façam parte do documento, considera-se necessário retomá-los para melhor entendimento do capítulo. Em destaque:

- *Gestão do Conhecimento*: possui a finalidade de entender, focar e gerir, de forma sistemática, bem como explicitar para deliberar sobre a construção de conhecimento, sua reutilização, renovação e aplicação (WIIG, 1997).
- *Referencial semântico*: objetiva a representação formal de conhecimento com suas complexas estruturas de relações (LEGG, 2007).
- *Ontologia*: é uma especificação de conhecimento consensual sobre um modelo abstrato de domínio, definida explicitamente em termos de conceitos, suas propriedades e relações por meio de axiomas, possibilitando, assim, que seja automaticamente processável por máquinas (STUDER et al., 1998).
- *Relações semânticas*: uma relação semântica trata-se de combinações entre conceitos, entidades. Obrigatoriamente as relações semânticas apresentam conceitos/entidades como parte integral da relação e devem conferir significado para o contexto semântico (KHOO e NA, 2006). Rector e Yunes (1980) consideram o contexto semântico como fator predominante no qual as relações semânticas situam-se como relações associativas que ocorrem dentro deste contexto onde os termos têm sentido e função.
- *Base de Conhecimento*: A base do conhecimento não é uma simples coleção de informações. A tradicional base de dados com dados, arquivos, registros e seus relacionamentos estáticos são aqui substituídos por uma base de regras e fatos e também heurísticas que correspondem ao conhecimento do especialista, ou dos especialistas do domínio sobre o qual foi construído o sistema (NILSON, 1982).

- *Vetor Semântico*: são geralmente baseados em uma matriz de co-ocorrência, uma forma de representar os seguintes valores – a frequência com que ocorrem palavras; os pesos que estas palavras representam em determinado contexto; e outras características que a linguística permite. Vetor semântico é uma matriz de termos de uma fonte linguística (e.g. documentos, ontologia) com representação matemática o que permite aplicações estatísticas e probabilísticas (JURAFSKY e MARTIN, 2015).
- Modelo de espaço vetorial (*Vector-Space Model* - VSM) (SALTON et al., 1975) é um modelo matemático que pode ser definido como uma estrutura lógica consistente projetada para corresponder a alguma entidade física, biológica, social, psicológica ou conceitual. VSM foi desenvolvido para o sistema SMART IR por Gerard Salton e colaboradores (1975). O sistema SMART foi pioneiro no desenvolvimento de muitos dos conceitos que ainda estão em uso nos motores de busca modernos (MANNING, 2009) (DUBIN, 2004).
- *Reconhecimento de padrões*: é definido como o processo pelo qual um padrão recebido é atribuído a uma classe dentre um número pré-determinado de classes (HAYKIN, 2001)
- *Sistemas Especialistas*: são tradicionalmente definidos como programas computacionais que modelam o conhecimento e emulam o processo de raciocínio de um especialista humano na resolução de um problema específico de um determinado domínio (TAYLOR e LUBKEMAN, 1989)(SUSTAETA et al, 1989)(LAVALLE e RODRIGUEZ, 1989).
- *Extração de características de Haralick*: descrevem uma metodologia para descrição da textura da imagem, onde são definidas diversas características advindas do cálculo de matrizes de co-ocorrência, que são matrizes que contam as ocorrências de níveis de cinza em uma imagem. Essas características servem como medida para a diferenciação de texturas que não seguem um determinado padrão de repetitividade, fornecendo informações relevantes para classificação das mesmas (HARALICK et al., 1973).
- *Redes Neurais Artificiais*: também conhecidas como métodos conexionistas, RNAs são inspiradas nos estudos da maneira de como se organiza e como funciona o cérebro humano. Trata-se de modelos computacionais não lineares, inspirados na estrutura e na operação do cérebro humano, que procuram reproduzir características humanas, tais como: *aprendizado, associação, generalização e abstração*. Redes Neurais são

efetivas no aprendizado de padrões a partir de dados não lineares, incompletos, com ruído ou compostos de exemplos contraditórios (PACHECO, 2002).

- UML (do inglês: *Unified Modeling Language*). Segundo Booch (1999), trata-se de linguagem gráfica de modelagem designada para especificar, visualizar, construir e documentar artefatos de um sistema de software intensivo. A UML foi adotada como notação padrão em projetos de desenvolvimento de software pelo Grupo de Gerenciamento de Objetos (OMG), e caracteriza-se como o principal protocolo de comunicação da indústria de software.

5.2 Modelo conceitual da tese

5.2.1 Visão

O modelo conceitual apresentado neste trabalho visa a especificação dos elementos estruturais da pesquisa, permitindo assim mensurar o escopo do referencial semântico, assim como a base teórica do domínio do conhecimento.

O modelo conceitual que suporta esta pesquisa sustenta-se nos seguintes elementos estruturais fundamentais: Conhecimento, Semântica, Padrão e Ator (Figura 28). *Ator* representa o grupo de profissionais que trabalha no contexto da problemática. As experiências e habilidades, oriundas dos Atores, são considerados imprescindíveis para a formalização do conhecimento. Este conhecimento, formalizado, cria um arcabouço *Semântico* que é compartilhado entre as entidades e os atores envolvidos no domínio do problema.

O domínio do problema apresenta *Padrões*, representados por conceitos, relações, imagens ou outras formas de comunicação, que podem ser identificados por técnicas de reconhecimento de padrões.

O modelo conceitual apresenta um contexto dinâmico, visto que a entidade *conhecimento* situa-se como elemento estrutural mediador entre os demais elementos.

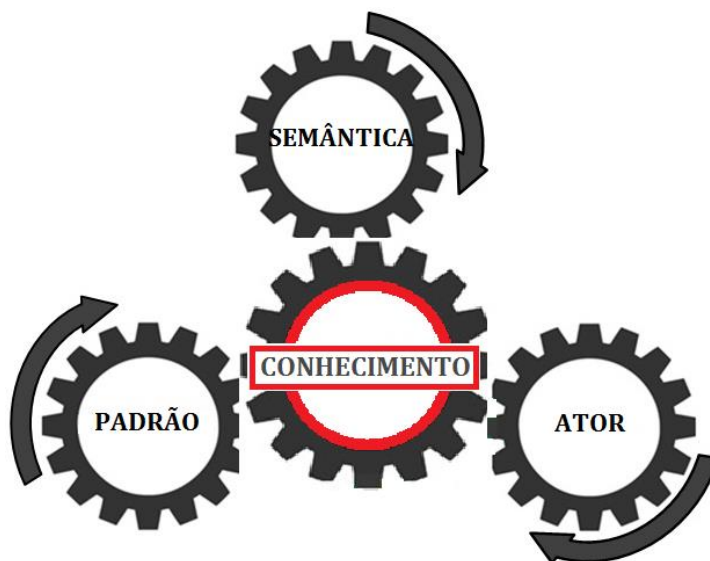


Figura 28. O Modelo Conceitual.

5.2.2 Instanciação do Modelo

A instanciação do modelo conceitual que suporta esta pesquisa apresenta os elementos estruturais fundamentais aplicados no âmbito da botânica (Figura 29). O *Ator* representa o grupo de profissionais que trabalha no âmbito da botânica (mais especificamente no processo de identificação botânica), nomeadamente botânicos, taxonomistas, mateiros (parataxonomistas) e engenheiros florestais. As experiências e habilidades (dos atores) usadas e necessárias na identificação botânica, servem para a formalização do conhecimento.

O domínio da botânica apresenta *Padrões* e, neste trabalho, o Referencial Semântico torna-se um instrumento apropriado para mapeá-los, visto que a estrutura semântica desenvolvida neste trabalho utiliza conceitos, relações, imagens ou outras formas de comunicação, que podem ser identificados por técnicas de reconhecimento de padrões.

O domínio da botânica apresenta um contexto dinâmico, visto que novos conhecimentos são continuamente inseridos no domínio. Desta forma, na instanciação do modelo conceitual a entidade *conhecimento* interage com as demais entidades atuando como fomentador da estrutura semântica.

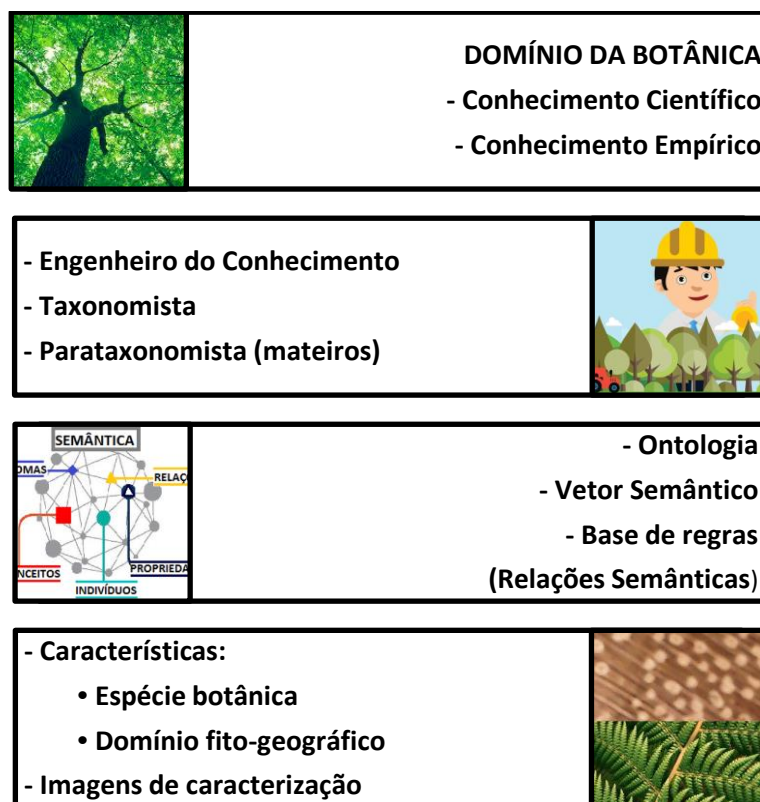


Figura 29. Instanciação do Modelo Conceitual

5.2 O Projeto

O projeto tem finalidade especificar e representar de forma organizada o que está sendo desenvolvido nesta tese. Para tal é necessário definir o escopo do projeto e a utilização dos diagramas. Vale ressaltar que uma parte do conteúdo será implementado utilizando linguagens computacionais e que a outra parte será implementado utilizando ferramentas computacionais.

O escopo do projeto abrange todos os elementos que estruturam o modelo conceitual desta tese, nomeadamente o referencial semântico, e seus instrumentos e cenários de validação (Figura 30). O referencial semântico é composto pela ontologia Onto-AmazonTimber e por vetores semânticos. Como instrumento de validação foram desenvolvidos e utilizados dois sistemas especialistas: um para classificação de imagem de madeira e o outro para identificação botânica – Inventário florestal.

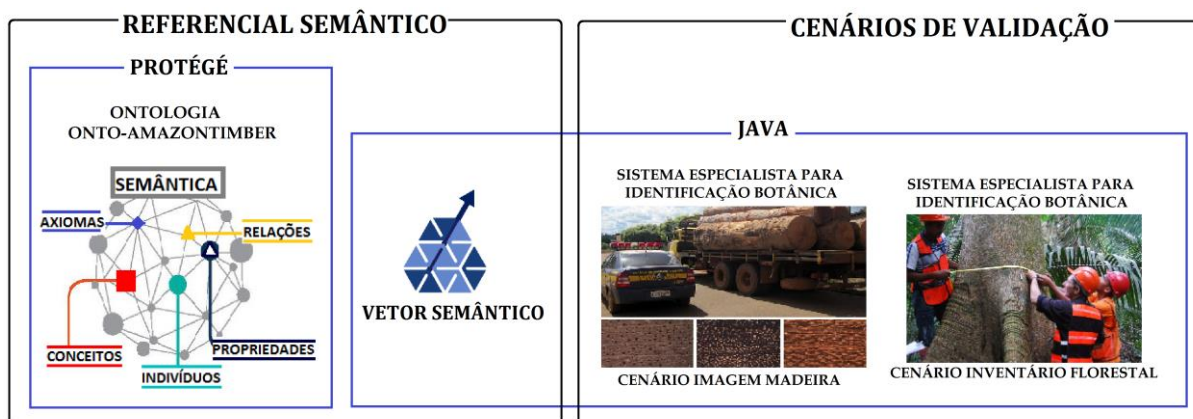


Figura 30. Visão geral do Projeto

Este capítulo não apresenta de forma detalhada todos os diagramas que foram trabalhados neste projeto. Foram escolhidos os diagramas que apresentam maior relevância para o entendimento deste trabalho.

5.2.1 Processo de criação do Referencial Semântico

O processo de criação do Referencial Semântico (Figura 31) inicia-se com a aquisição de conhecimento. O engenheiro do conhecimento tem como função abstrair o conhecimento de especialista da área do domínio problema, assim como identificar e rotular outras fontes de conhecimento. Após a aquisição, o conhecimento deve ser formalizado (etapa 2) e formalizado na Onto-AmazonTimber, que é composta por conceitos, indivíduos, propriedades, axiomas e relações semânticas, permitindo que o conhecimento seja utilizado em aplicações diversas.

As relações semânticas representam o cerne da Onto-AmazonTimber, e expressam a interação entre conceitos e indivíduos regidos por axiomas com funções de restritivas e extensivas. Ocorre que cada relação apresenta um valor semântico podendo ser mensurada. Desta forma, utiliza-se nesta pesquisa o vetor semântico (etapa 3) para mensurar o grau de relevância de relações semânticas, o que permite criar regras e qualificá-las quanto à importância, além de aumentar a acurácia dos resultados em diagnósticos e classificações.

Para evitar a degradação da Onto-AmazonTimber, é pertinente a constante atualização. Para tal, é necessária classificar (etapa 4) novas fontes de conhecimento e alocá-las em categorias pré-definidas. Além da prática de atualização, é necessário tanto uma constante avaliação (etapa 5) dos resultados de interações semânticas quanto as correções em caso de conflitos.

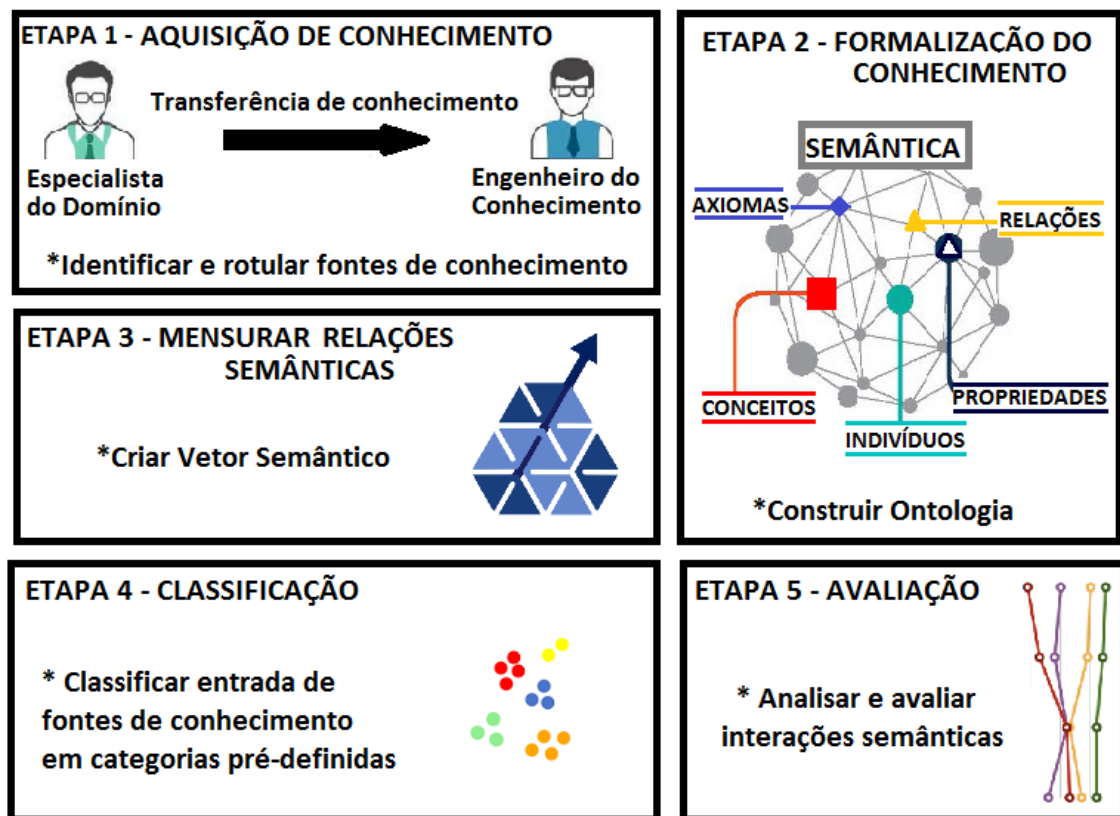


Figura 31. Processo de desenvolvimento da estrutura semântica

5.2.2 Especificação de Requisitos

Esta etapa envolve atividades que determinam o objetivo do sistema e todas as restrições associadas a ele, neste trabalho, utiliza-se como notação a UML para especificação de requisitos. Como definição de escopo a especificação de requisitos abrange o referencial semântico, o sistema especialista para classificação de imagem de madeira e o sistema especialista para identificação botânica – Inventário florestal.

Neste contexto, a especificação de requisitos é descrito em três visões diferentes - funcional, arquitetural e comportamental). A visão funcional mostra as interações entre entidades externas e o referencial semântico. A visão arquitetural apresenta, em um modelo de 3 camadas (camadas de apresentação, controle e dados), a estrutura que suporta o referencial semântico. A visão comportamental descreve a interação entre os componentes estruturais deste trabalho.

5.2.2.1 Visão Funcional

Casos de uso especificam o comportamento do sistema ou parte(s) dele e descrevem a funcionalidade do sistema desempenhada pelos atores. Um ator representa uma entidade que interage com o sistema e, apesar de ser externo a ele, está inserido em seu ambiente. Os casos de uso apresentam a sequência específica de ações e interações entre atores e o sistema, por sua vez os relacionamentos integram os elementos, isto inclui: associações com casos de uso e generalização / especialização com outros atores; o relacionamento de inclusão (<<include>>) deve ser sempre direcionado do caso de uso base para o caso de uso que é incluído; o relacionamento de extensão (<<extend>>) deve ser sempre direcionado do caso de uso de extensão para o caso de uso estendido (base).

O diagrama de caso de uso – Atribuições dos atores (Figura 32) envolve dois atores diferentes, o especialista do domínio e o engenheiro do conhecimento, no qual participam dos seguintes casos de uso:

- Conceitualizar ontologia: definir e contextualizar o domínio do conhecimento.
- Coletar fontes de conhecimento: este caso de uso está relacionado à tarefa de coleta de fontes de conhecimento relevantes que serão usadas para avaliar a prova de conceito.
- Manter ontologia: atividades relacionadas a adicionar, atualizar e deletar componentes da ontologia.
- Formalizar ontologia: está relacionada à tarefa de especificar formalmente a ontologia de domínio em OWL.

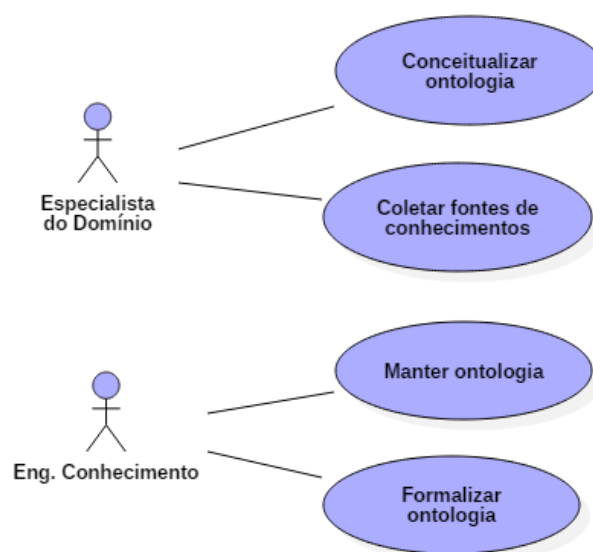


Figura 32. Diagrama de Caso de Uso - Atribuições dos atores

O engenheiro do conhecimento tem como função formalizar o conhecimento com a criação do referencial semântico. Desta forma, o diagrama de caso de uso – Gestão do conhecimento (Figura 33) reporta as ações necessárias para a implantação e manutenção do referencial semântico. Dentre os casos de uso destacam-se:

- “Manter ontologia”: atividades relacionadas a adicionar, atualizar e deletar componentes da ontologia. Entende-se componente da ontologia como conceitos, instâncias, propriedade de objetos, propriedade de dados, relações semânticas e axiomas;
- “Manter vetor semântico”: atividade relacionada à criação do vetor semântico com atribuição de pesos em características que podem ser provenientes da pesquisa do usuário ou de cada espécie botânica contida na ontologia. Outra ação está relacionada à atividade de calcular a similaridade entre estes dois vetores semânticos (vetor da pesquisa do usuário X vetor das espécies botânicas).

O diagrama de caso do uso – Interface (Figura 34) apresenta as ações necessárias para o usuário identificar uma espécie botânica, podendo identificar por características ou por imagens da madeira. Vale ressaltar que o usuário pode ou não ser um especialista no domínio da botânica.

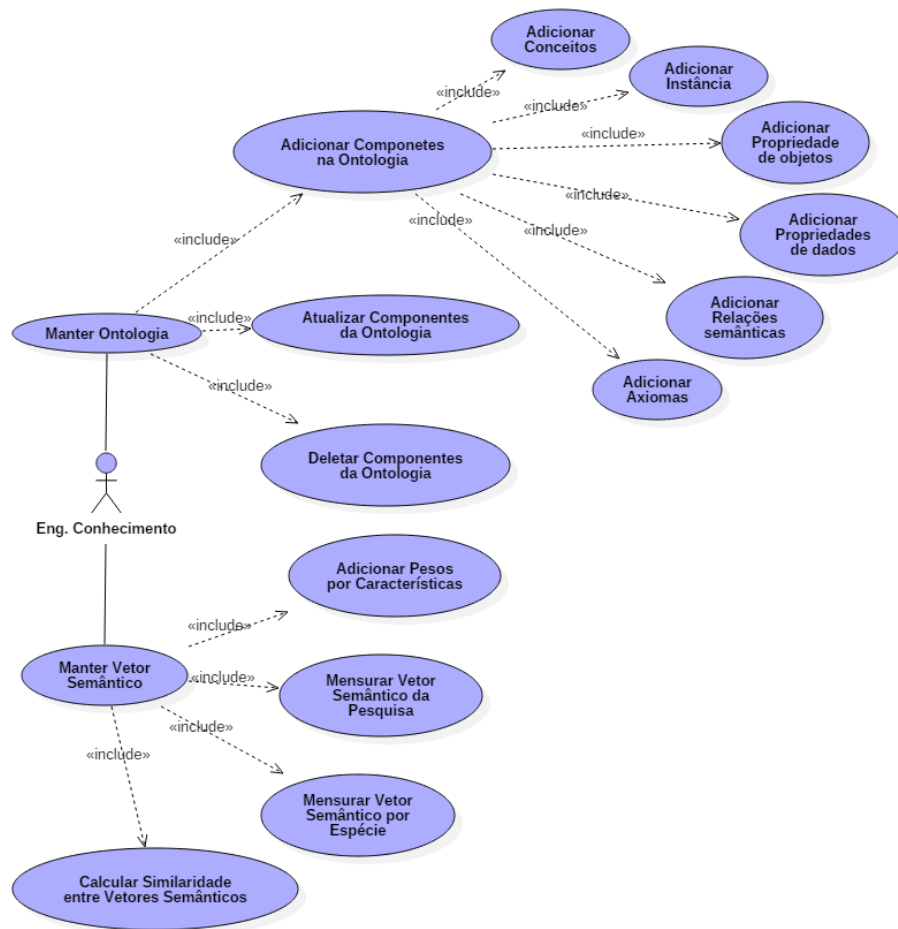


Figura 33. Diagrama de Caso de Uso – Gestão do conhecimento

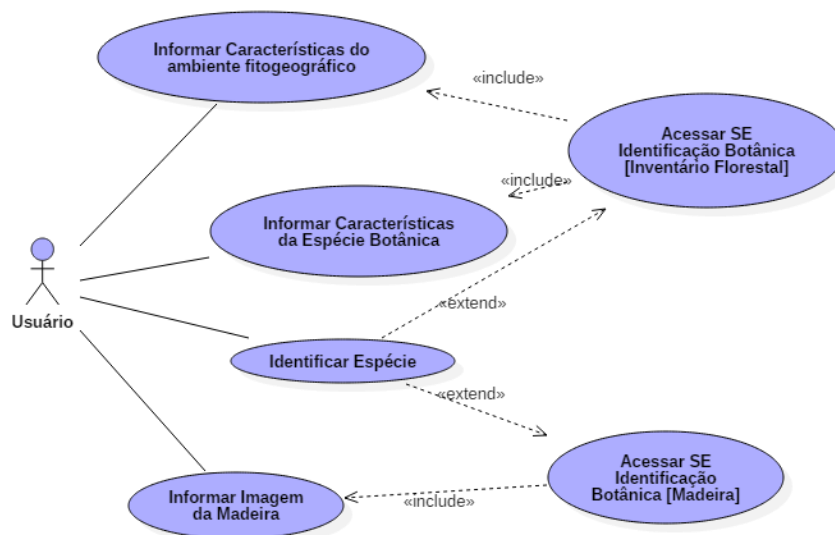


Figura 34. . Diagrama de caso do uso – Interface

5.2.2.2 Visão Arquitetural

A visão arquitetural do referencial semântico é apresentada através do diagrama de componentes da UML, que tem como função particionar um sistema em blocos para melhor entendimento e organização. O diagrama de componentes (Figura 35) cobre três camadas diferentes: Apresentação, Controle e Dados. A camada de apresentação contém os componentes que lidam com a interação entre agentes externos e o sistema - os componentes da interface. A camada de controle mantém componentes responsáveis pela implementação da lógica de negócios para executar as operações necessárias para operar a ontologia, os sistemas especialistas para identificação botânica em imagens de madeira e em nível de inventário florestal. Finalmente, a camada de dados contém componentes para manipular todas as fontes de dados usadas no referencial semântico.

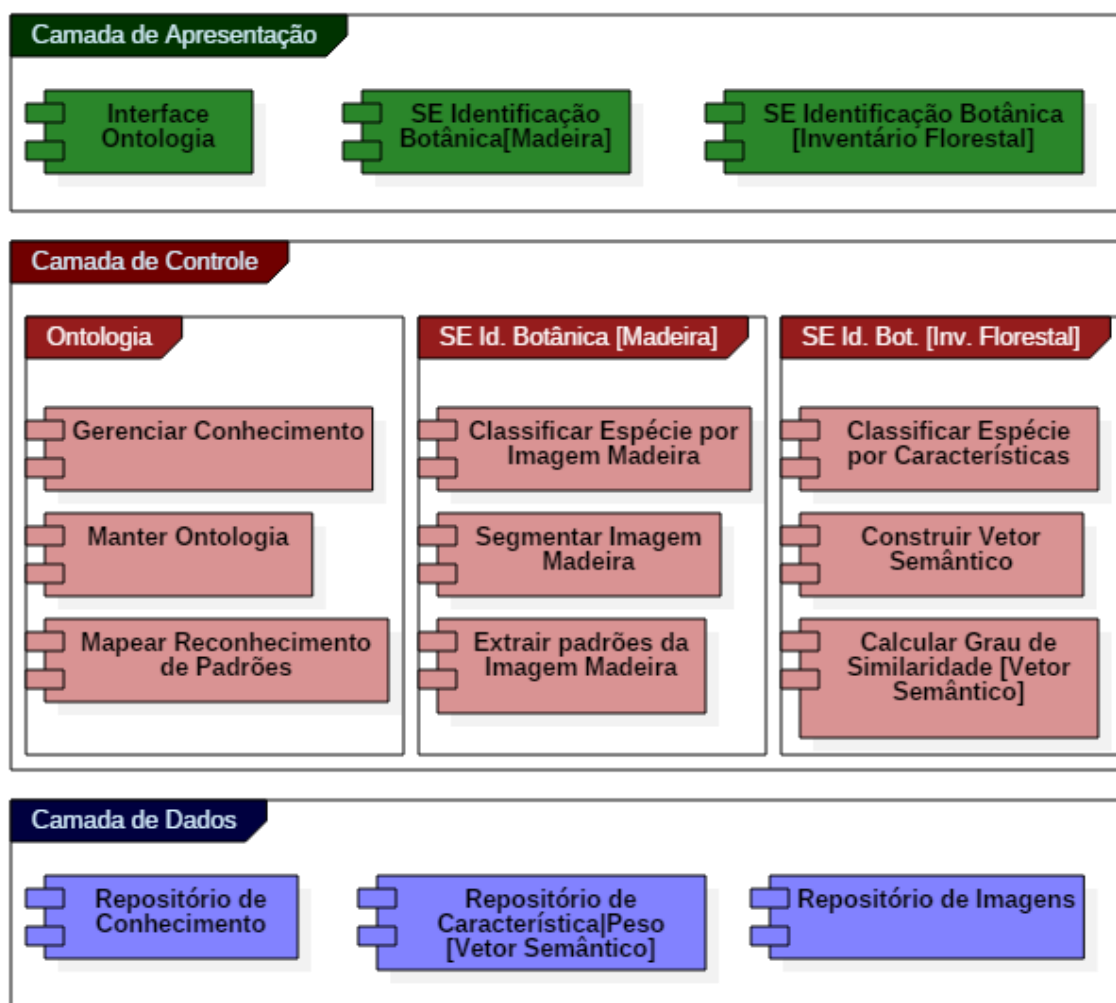


Figura 35. Diagrama de Componentes

5.2.2.3 Visão Comportamental

Formalmente, o diagrama de sequência da UML tem como função representar o comportamento de vários objetos dentro de um contexto a partir das mensagens que são trocadas entre eles. Nesta tese, este diagrama é usado de maneira mais ampla, para representar as interações entre os vários componentes desenvolvidos.

O diagrama de sequência – Inventário florestal (Figura 36) para descrever a sequência de ações necessárias para identificação de espécies botânica em nível de inventário florestal.

O diagrama de sequência se inicia com a ação de acesso do usuário ao sistema que se comunica com a ontologia que, comunica-se com um repositório de imagem. Após a interface carregada, o usuário informa as características botânicas encontradas da espécie que quer identificar. Por sua vez, o sistema especialista cria um vetor semântico das características elencadas pelo usuário, e ao mesmo tempo acessa os vetores semânticos das espécies já catalogadas, consultam-se os pesos de todas as características. De posse dos vetores semânticos, calcula-se a similaridade entre o vetor semântico das características elencadas pelo usuário com os vetores semânticos de todas as espécies catalogadas, retornando as espécies que apresentam maior similaridade com as características apresentadas pelo usuário.

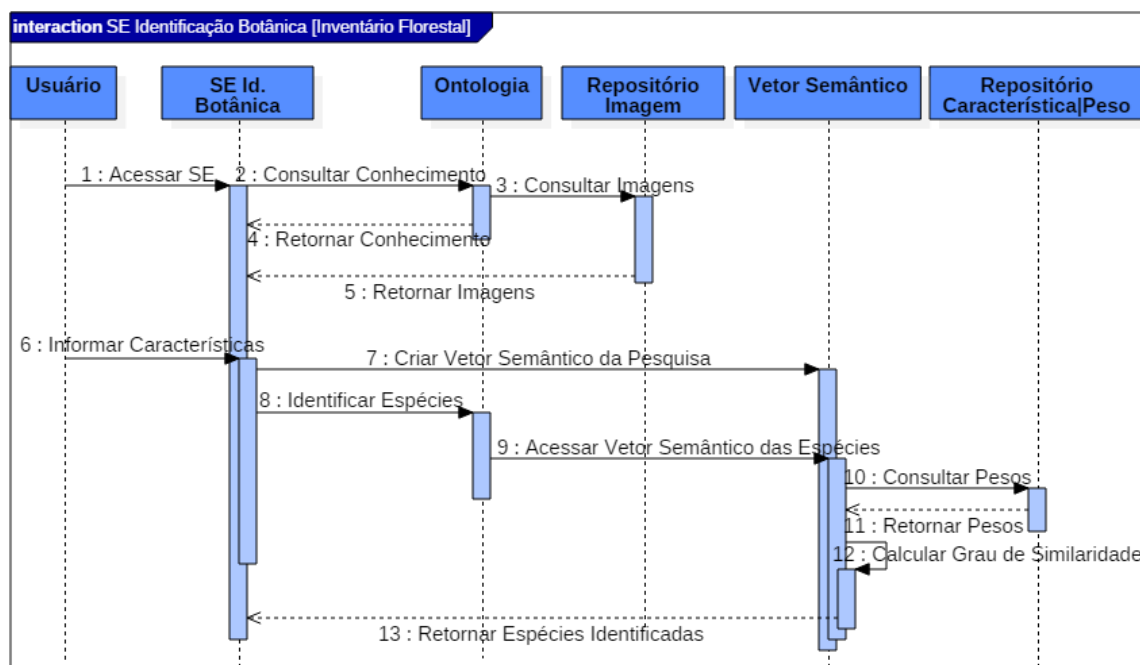


Figura 36. Diagrama de Sequência – Inventário florestal

O diagrama de sequência – Madeira (Figura 37) descreve a sequência de ações necessárias para identificação de espécies botânicas, utilizando imagens de madeira com suporte do referencial semântico.

O diagrama de sequência se inicia com uma rotina de ações responsáveis para calibrar o método de classificação, tal evento não apresenta regularidade, ocorre quando novas imagens são adicionadas ou há modificações de configuração do método de classificação. Após esta rotina, o usuário captura e informa a imagem da madeira para ser identificada e o sistema especialista segmenta a imagem para posterior extração de padrões da imagem. Tais padrões serão analisados e classificados, para que assim se possa identificar a espécie botânica à qual pertence a imagem da madeira.

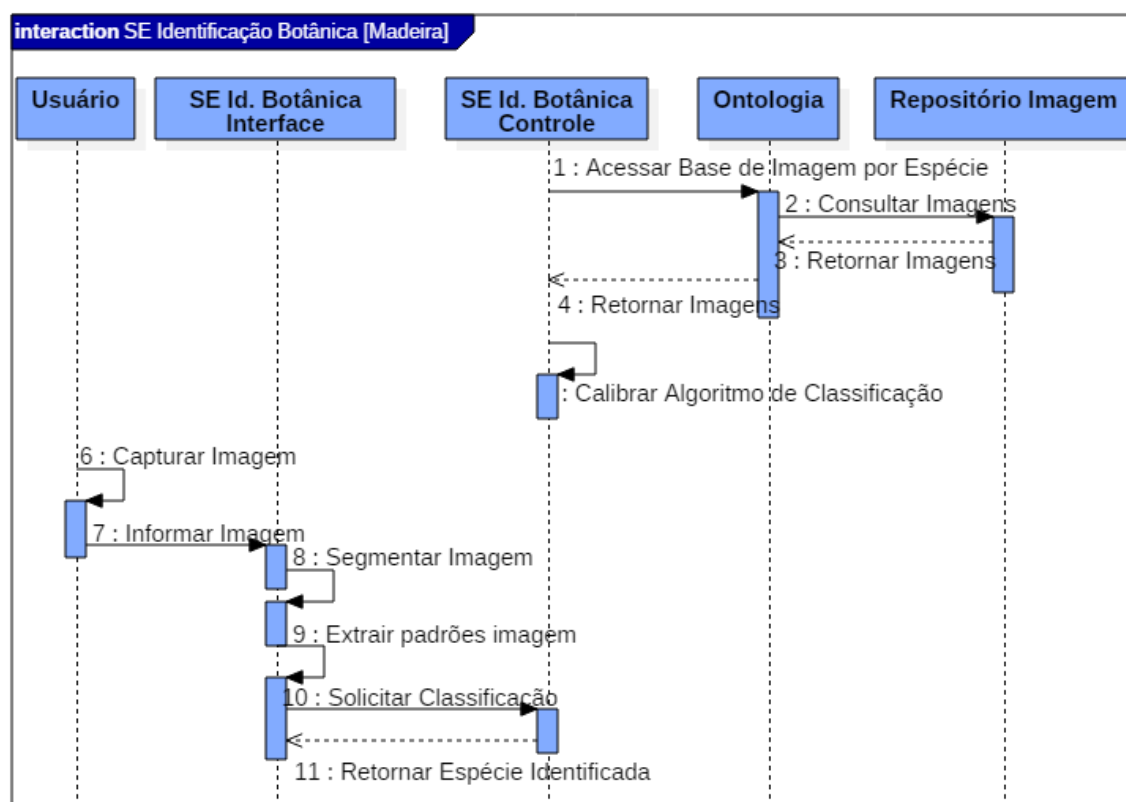


Figura 37. Diagrama de Sequência – Madeira

5.3 Implementação

Este tópico objetiva descrever como ocorre a implementação do referencial semântico e seus cenários de validação, respectivamente codificados por ferramentas computacionais e

linguagem de programação. Além de reportar as tecnologias utilizadas e os produtos gerados no decorrer da implementação.

5.2.1 Tecnologias usadas na implementação

As tecnologias empregadas nesta pesquisa identificam os componentes utilizados na implementação do referencial semântico e dos sistemas especialistas utilizados na validação, assim como sua interação e relacionamento entre as partes integrantes.

Os cenários de validação do referencial semântico apresentam especificações tecnológicas distintas, entretanto compartilham um objetivo único de suporte ao processo de identificação de espécies botânicas da Amazônia, com intuito de validar o conhecimento e estrutura semântica do referencial semântico.

As tecnologias representadas na figura 38 apresentam uma estrutura que aporta o modelo conceitual da tese, possibilitando sua integração com tecnologias adjacentes, ou seja, integração com metodologias, métodos, ferramentas, e demais recursos computacionais e tecnológicos que saem do escopo do referencial semântico.

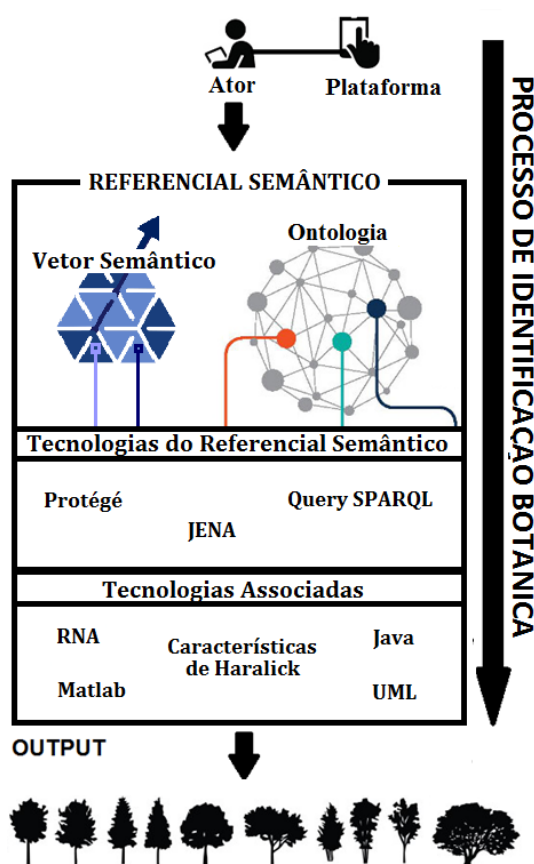


Figura 38. Tecnologias utilizadas

Neste trabalho utiliza-se a plataforma de comunicação por dispositivos móveis devido à característica de usabilidade, capacidade de capturar imagens, mobilidade e conectividade a uma maior distância dos centros urbanos, assim como flexibilidade no desenvolvimento aplicado a imagens e bases de conhecimento.

Por conseguinte, o ator (representando os profissionais que trabalham no âmbito da botânica) interage com os conhecimentos oriundos da ontologia representados por relações semânticas compostas por características e imagens botânicas. A ontologia neste trabalho situa-se como alicerce da tecnologia semântica, o que permite o usufruto por tecnologias associadas oriundas de outros arranjos e domínios do conhecimento.

As tecnologias do referencial semântico permitem a sua implementação; por sua vez, as tecnologias associadas amplificam os serviços e funcionalidades do referencial semântico com recursos computacionais que apóiem o processo de identificação.

5.2.2 A Ontologia ONTO-AMAZONTIMBER

A ontologia representa uma parte do domínio de conhecimento da Botânica, mais especificamente focada no processo de identificação taxonômica, apresentando um acervo de características e particularidades de uma vasta quantidade de espécies florestais da Amazônia.

Onto-AmazonTimber é uma ontologia de aplicação posto que se trata de uma ontologia no domínio da botânica que será utilizada no processo de identificação de espécies da Amazônia. O conhecimento científico necessário para sua construção foi extraído das seguintes fontes: literatura relevante, entrevista com taxonomistas, e identificação por características técnicas e concisas. Por sua vez, o conhecimento empírico foi capturado por meio de atividades práticas com percepção de características experimentais, através da experiência dos mateiros e dos taxonomistas.

Com pode-se observar na Figura 39, outras fontes de conhecimento utilizadas são: Sistema computacional com características botânicas de espécies do Brasil desenvolvido pelo Laboratório de Produtos Florestais (LPF) do Serviço Florestal Brasileiro (SFB); um banco de dados de espécies botânicas do Brasil, o que permitirá acessar um grande volume de características e aspectos das espécies botânicas, incluindo nomenclaturas vernaculares por região. Este banco de dados, gerenciado pelo SFB, foi desenvolvido conjuntamente pelo Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA) e pela *International Tropical Timber Organization* (ITTO).

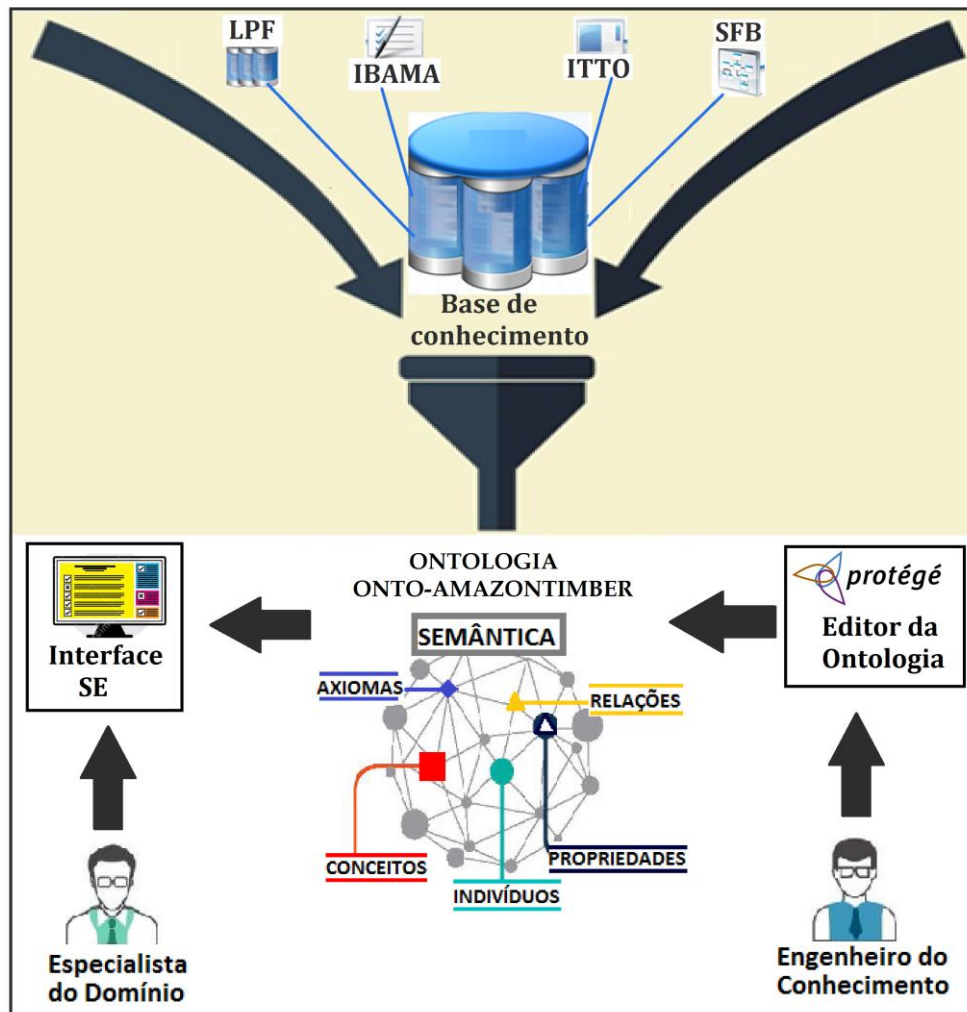


Figura 39. Base de conhecimento do domínio

5.2.2.1 A ESTRUTURA TAXONÔMICA

A estrutura taxonômica da ontologia (Figura 40) divide-se em: *External Characteristic* (Características Externas), *Genus* (Gênero -hierarquia das espécies), *Structural Characteristic* (Características Estruturais). Tal estrutura possibilita navegar de forma intuitiva detalhes das características das espécies botânicas. Vale ressaltar a importância da entidade *Specie* no contexto semântico da ontologia, visto que as demais entidades estão estruturadas para referenciar as instâncias da entidade *Specie*, ou seja, grande parte das entidades define alguma característica de uma espécie botânica.

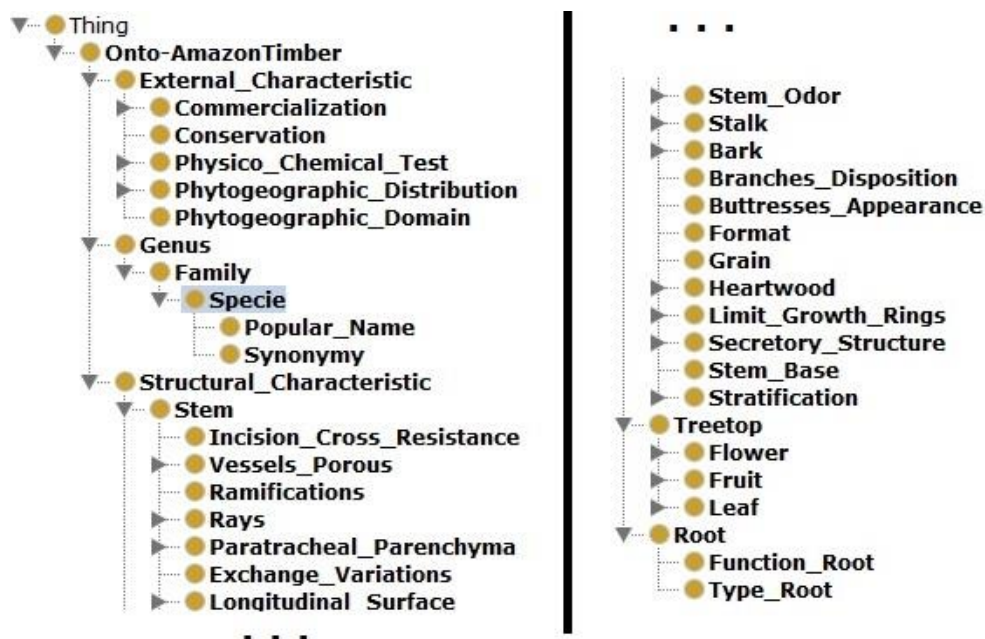


Figura 40. Estrutura taxonômica da ontologia

A Onto-AmazonTimber foi construída em duas línguas: inglês e português (Figura 41). Esta característica bilíngue é suportada pela definição de *labels* nas entidades e nos indivíduos, permitindo que sejam acessados e dispostos conforme a opção do idioma. A disponibilidade de idiomas situa-se como parte determinante na usabilidade da ontologia, visto que permite o acesso ao conhecimento e interação da comunidade científica com potenciais usuários (e.g. mateiro, polícia ambiental, entre outros).

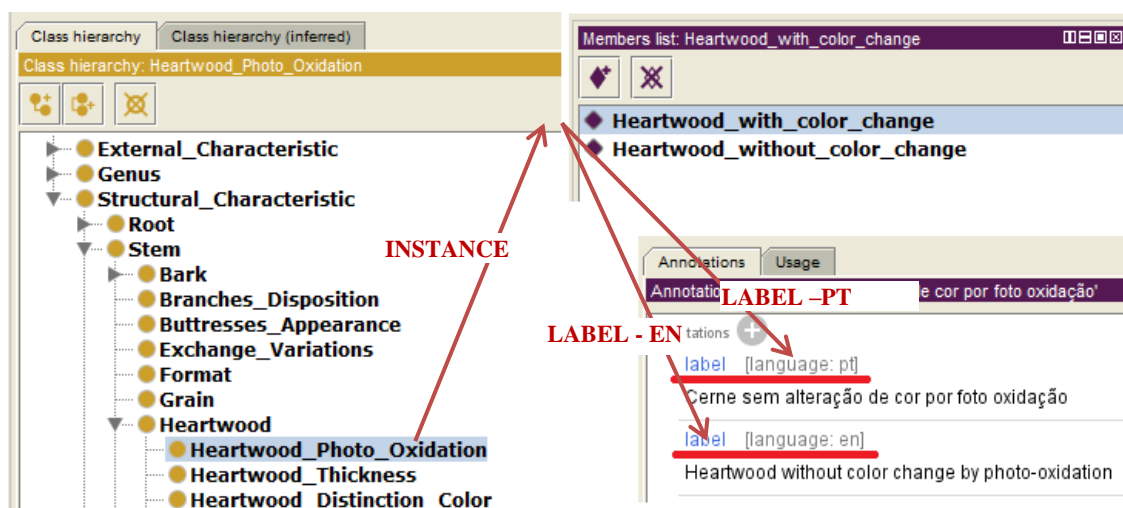


Figura 41. Ontologia – Label inglês/português

5.2.2.2 ESTRUTURA SEMÂNTICA E AXIOMAS

A Onto-AmazonTimber contém estrutura semântica composta por conceitos, relações, propriedades e axiomas, instanciados e inter-relacionados para compor o domínio de conhecimento em questão. Atualmente, a Onto-AmazonTimber possui 180 entidades, 2427 relações, 9 propriedade de objetos, 3 propriedades de dados, 4619 axiomas lógicos e 869 indivíduos.

As relações semânticas foram sendo delineadas e validadas concomitantemente ao processo de definição das entidades, visto que a profundidade do conhecimento modelado na Ontologia depende da experiência adquirida nas atividades do processo de identificação de espécies no âmbito da botânica. Contudo, tais relações semânticas não se limitam à estrita prática de identificação botânica, posto que devem evidenciar em sua completude a instanciação do modelo semântico desenvolvido neste trabalho. Deste modo, as relações semânticas propiciam uma proposta de integração de conhecimento entre os mateiros e taxonomista, norteando as equiparações entre nomes vernaculares e científicos.

As relações semânticas, em geral, são formadas por expressões regulares que se dividem em *propriedade de objetos e propriedade de dados*. A propriedade de objetos tem como função a comunicação entre as entidades, propiciando relações complexas que destacam-se como *hot spot* na construção do conhecimento. Por sua vez, a propriedade de dados tem como função conectar uma entidade a um dado literal, ou seja, permitem assim armazenar um volume dados referentes às entidades.

A Onto-AmazonTimber apresenta como propriedade de objetos as seguintes expressões: *Classified by*, *Composed by*, *Formed by*, *Included in*, *It has*; *It has popular name*, *It has scope*, *It has synonymy* e *It has family*. Tais propriedades apresentam características que permitem especificar suas particularidades. Dentre as características apresentadas nas propriedades desta ontologia destacam-se a assimetria (ausência de relação de fluxo inverso com a mesma nomenclatura), não funcional (as relações podem apresentar multiplicidade) e irreflexível (relações com único fluxo).

Como propriedade de dados, esta ontologia apresenta as seguintes expressões: *Has image*, *It has function* e *It has measured*. Tais expressões regulares juntamente com entidade e dados literais compõem as relações semânticas que delineiam os axiomas estruturados, que buscam evidenciar particularidades de entidades.

Destaca-se a importância da propriedade de dados *Has image* que permite criar uma relação semântica entre entidades e imagens, possibilitando desta forma a construção de uma base de imagens no âmbito da botânica integrada ao referencial semântico.

Axiomas são utilizados para modelar sentenças consideradas sempre verdadeiras. São restrições aplicadas à estrutura semântica da ontologia no intuito de formalizar sentenças e regras estipuladas pelos especialistas do domínio do conhecimento.

Desta forma, destaca-se o axioma com maior valor semântico da ontologia o qual descreve as características de cada espécie botânica como se pode observar na figura 42. A entidade *Specie* contém várias instâncias, dentre elas a *Dipteryx odorata* que, por sua vez, contém várias relações semânticas de restrição que se dividem por propriedades de objetos definidos por (*Dipteryx odorata included in list most marketed*, *Dipteryx odorata it has popular name Cumaru*) e por propriedades de dados definidos por (*Dipteryx odorata has image "c:\\imagemontologia/..."*). Tais axiomas possibilitam inferir qual a espécie botânica pertence às características explicitadas por estas relações restrição.

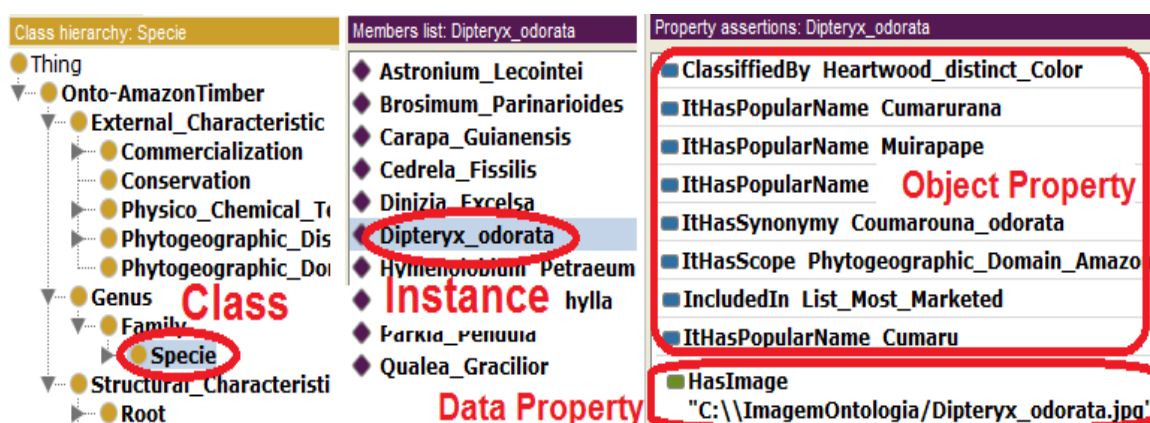


Figura 42. Instanciação do Modelo Conceitual

Contudo outras formas de axiomas podem ser identificadas na ontologia. Na construção das entidades (Classes), pode-se restringir as instâncias de uma determinada classe, com a aplicação do axioma *Disjoint*. Desta forma, quando classes são *Disjoint*, um indivíduo (ou objeto) não poderá ser instância de mais de uma dentre as classes do grupamento *Disjoint*, ou seja, entre as classes *Disjoint* apenas uma poderá ser instanciada. Pode-se observar na Figura 43 a relação *Disjoint* entre as entidades “Dry”, “Seed” e “Meaty”, tornando obrigatória uma única representação para *Fruit*.

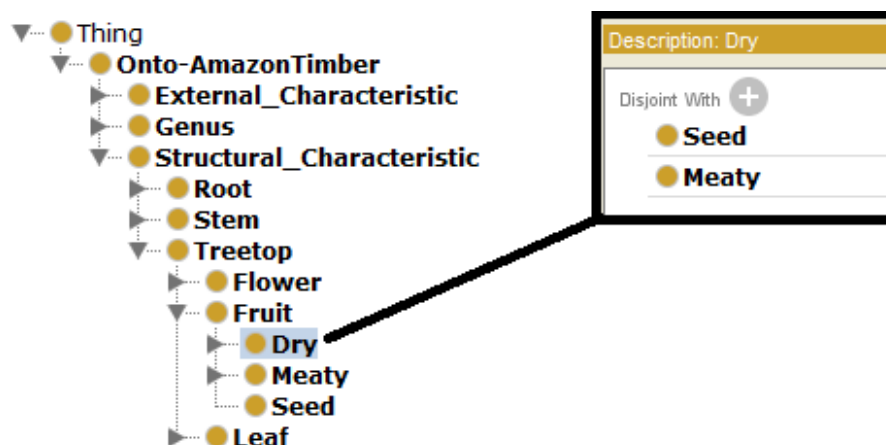


Figura 43. Disjoint Class

5.2.3 Sistema Especialista para identificação botânica – Inventário Florestal

Embora os mateiros apresentem um grande conhecimento empírico sobre a floresta, pouco científico e, no domínio da botânica, é geralmente fraco, o que dificulta a identificação botânica e a correta nomenclatura das espécies (cientificamente falando). Assim, torna-se necessário auxiliá-los nos processos de tomada de decisão de identificação botânica. Desta forma, o cenário do inventário florestal propõe um sistema computacional que identifica espécies botânicas, baseado em características observadas na espécie e no ambiente em que estão inseridas, com auxílio de imagens que tipificam as características.

O cenário permite ao usuário inserir características observadas da espécie botânica no momento do inventário florestal, obtendo, como resposta, as espécies e seus respectivos nomes vernaculares. Contudo quanto maior o número de características identificadas pelo usuário maior será a precisão da identificação botânica.

5.2.3.1 BASE DE CONHECIMENTO

A base de conhecimento do sistema especialista proposto situa-se como parte integrante de um referencial semântico, formalizado pela ontologia Onto-AmazonTimber.

O conhecimento científico necessário para sua construção foi extraído das seguintes fontes: literatura científica, entrevista com taxonomistas, base de imagens, base de regras. Por sua vez, o conhecimento empírico foi capturado por meio de atividades práticas com percepção de características experimentais, através da experiência dos mateiros e dos taxonomistas. Outras fontes de conhecimento utilizadas são: Sistema computacional com características botânicas de espécies do Brasil desenvolvido pelo LPF; um banco de dados de

espécies botânicas do Brasil, o que permitirá acessar um grande volume de características e aspectos das espécies botânicas, incluindo nomenclaturas populares por região. Este banco de dados, gerenciado pelo SFB, foi desenvolvido conjuntamente pelo IBAMA e pelo ITTO.

Angariou-se, nesta pesquisa, um banco de imagens macroscópicas do cerne de madeiras, coletado no laboratório de madeira da Universidade Federal do Oeste do Pará (UFOPA). Tais imagens permitem evidenciar padrões de comportamentos da imagem como características cor (pigmentação), geometria e frequência dos raios (parênquimas), vasos, poros, traqueoides, xilemas, entre outros conjuntos de primitivas.

As regras e heurísticas dos especialistas são formalizadas na ontologia *Onto-AmazonTimber* por relações semânticas que compõem restrições axiomáticas e interações com outras entidades que contemplam o referencial semântico. Vale ressaltar que relações semânticas não se limitam à estrita prática de identificação botânica, devem evidenciar em sua completude da instanciação do referencial semântico, propiciando assim uma proposta de integração de conhecimento entre os mateiros e taxonomista, norteando as equiparações entre nomes populares e científicos.

5.2.3.2 INTERAÇÃO COM A ONTOLOGIA – JENA

O JENA é um framework Java que permite trabalhar em ambiente de programação com manipulação dinâmica de modelos RDF, representadas pelos recursos, propriedades e literals, formando as tuplas (predicate, [subject],[object]) que originam os objetos criados pelo java. O JENA apresenta um conjunto de funcionalidades para apoiar o desenvolvimento de aplicações no contexto de ontologias, além das funcionalidades para manipulação da linguagem OWL e uso do *Simple Protocol And Rdf Query Language* (SPARQL) (APACHE, 2015).

O JENA dispõe de um conjunto de métodos que permitem acessar os elementos de uma Ontologia (classes, propriedades e indivíduos), podemos utilizar os métodos iniciados com *list* como *listClasses()*, *listIndividuals()* ou *listSubClasses()*, como se pode observar respectivamente na tabela 1. Além disso, para identificar qual classe ou instância está sendo manipulada dentro das iterações, temos dois métodos básicos: *getURI()*, que retorna o nome completo ou a URI (prefixo + nome) do objeto; e *getLocalName()*, que retorna apenas o nome do objeto em questão.

Tabela 10. Métodos List() da API JENA

MÉTODOS	CÓDIGO
Listar Classes	<pre>OntClass concept : newM.listClasses().toList() concept.getLocalName()</pre>
Listar Instâncias	<pre>Individual intance:newM.listIndividuals().toList() intance.getLocalName()</pre>
Listar Sub-classes	<pre>OntClass concept = (OntClass) newM.getOntClass(uri); OntClass subConcept : concept.listSubClasses().toList() subConcept.getLocalName()</pre>

Outros métodos permitem uma maior especificação quanto ao acesso à estrutura ontológica. Por exemplo, o método *getObjectsFromObjectTriple* possibilita listar um conjunto de objetos da classe A que se relacionam através de uma propriedade específica com um outro objeto da classe B. Para melhor entendimento, visualiza-se o código a seguir:

```
OntologyInteraction listaCaracteristica = new
OntologyInteraction() ;

ArrayList<String>      objects      =      listaCaracteristica.
getObjectsFromObjectTriple("Dipteryx_odorata",
"ClassificadoPor");
```

Tal método *getObjectsFromObjectTriple* tem como função listar as características botânicas da espécie botânica “Dipteryx_odorata” que estão interligadas pela propriedade “ClassificadaPor”, tal relação semântica e podem ser observadas na figura 44. A classe “Species” tem uma série de objetos dentre estes o Dipteryx_odorata, por sua vez apresenta algumas propriedades de objetos que criam relações com outros objetos, como por exemplo *Heartwood_Distinct_Color*, instância da classe *Heartwood_Color*.

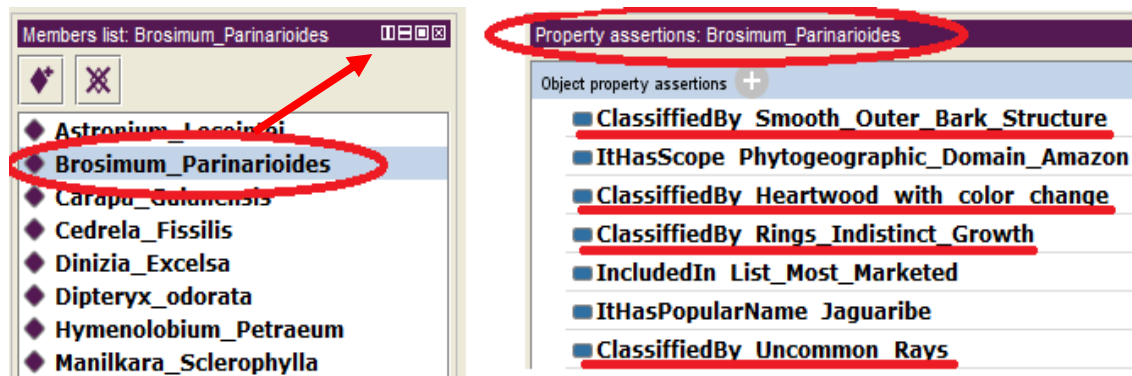


Figura 44. Relações semânticas obtidas pelo método *getObjectsFromObjectTriple*

O RDF permite interoperar, inferir e publicar dados, permite ainda a exploração de dados e associações, utilizando-se da linguagem de consulta SPARQL.

5.2.3.3 INTERAÇÃO COM A ONTOLOGIA -QUERY SPARQL

A partir da modelagem semântica apresentada na ontologia Onto-AmazonTimber, algumas consultas SPARQL foram implementadas com seus respectivos resultados. A query SPARQL mais relevante deste trabalho pode ser visualizada no código mostrado a seguir, o qual visa identificar uma espécie botânica, utilizando como critério algumas características informadas pelo usuário.

```
PREFIX rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
PREFIX mm:<http://www.semanticweb.org/2016/ontology8#>

SELECT DISTINCT ?Especie WHERE {

    ?Species mm:IncludedIn mm:List_Most_Marketed .

    ?Species mm:TemSinonimia mm:Coumarouna_odorata .

    ?species mm:TemNomeVernacular mm:Cumaru .

    ?Species mm:TemAbrangencia mm:Amazonia .

    ?Species mm:ÉClassificadaPor mm:Cener_Azulada.}
```

A consulta SPARQL contém um conjunto de padrões triplos chamados de um padrão de gráfico básico. Estes padrões triplos são como triplos RDF exceto com cada um dos sujeitos, predicados e objetos que podem ser variáveis (W3C, 2014).

A consulta é constituída por duas partes: a cláusula **SELECT** que identifica as variáveis selecionadas na consulta, como se pode observar o código 2 no estudo de caso, a variável “Especie” antecedida por um ponto de interrogação como sintaxe representativa de variável. Outro código observado trata-se do *Distinct* que tem como função excluir cláusulas repetidas na resposta. Outra cláusula **WHERE** fornece o padrão básico gráfico como critério de seleção, atuando como filtro restritivo, utilizando-se as relações semânticas existentes na variável de seleção. O padrão básico gráfico no estudo de caso é constituído por várias tuplas com padrão de três com uma única variável (?Species) na posição do objecto. Tais tuplas fazem referências às características como (List_Most_Marketed; Coumarouna_odorata; Cumaru; Amazônia; Cener_Azulada) selecionadas com critério para seleção da Species, integradas pelas respectivas propriedades de objetos (IncludedIn; TemSinonimia; TemNomeVernacular; TemAbrangência; ÉClassificadaPor).

As relações semânticas inferidas nesta query SPARQL podem ser observadas na figura 2, observando as particularidades e distinções dos métodos de acesso, os quais utilizam as características botânicas como variável para seleção no método *getObjectsFromObjectTriple* ou como critério de seleção para o método SPARQL.

5.2.3.4 O VETOR SEMÂNTICO

Os VSMs apresentam vantagens quanto à facilidade de manipulação usando álgebra linear, permitindo que um grau de semelhança entre vetores seja computado. A criação de vetores semânticos tem como base o VSM para a abordagem apresentada neste trabalho, visto que implementa a extração de conhecimento a partir de fontes de conhecimento em uma representação matricial, mais adequada para tratar a relevância entre os termos.

Neste trabalho, o sistema especialista utiliza o VSM para realizar representações de conhecimento tanto de fontes de conhecimento como de consultas, além de definir um formato comum entre essas representações, permitindo comparações entre elas. Mais especificamente, o sistema especialista para identificação botânica a nível de inventário florestal utiliza o algoritmo da Distância Euclidiana para aproximação entre vetores construídos das pesquisas pelo usuário e vetores semânticos construídos pelos termos da ontologia com incrementos de pesos por relevância.

Isto posto, identifica-se a problemática da especificação e generalização dos resultados no processo prévio de identificação botânica, para posterior proposta de solução, utilizando VSM, SV e Distância Euclidiana.

Análise dos Resultados Preliminares – Problemática da Generalização X Especificação

A produção de conhecimento no âmbito da botânica, mais especificamente no processo de identificação botânica, apresenta diferenciações quanto à disponibilidade de caracterização botânica que varia de espécie para espécie. Neste sentido, quanto mais pesquisada for uma espécie, mais características obtêm notoriedade científica. Tais variações de abrangência de conhecimento entre as espécies repercutem no referencial semântico Onto-amazonTimber.

As relações semânticas que caracterizam as espécies botânicas expostas no referencial semântico conduzem o processo de identificação botânica a fim de aumentar a acurácia dos resultados. Tais resultados sofrem influência quanto ao nível de profundidade das características apresentadas, o que está diretamente relacionado com o nível de experiência do usuário no processo de identificação botânica.

Neste sentido, caso o usuário forneça uma quantidade de informações genéricas e pouco específicas, o resultado da identificação botânica apresenta um elevado número de espécies, ocasionando o excesso de generalização dos resultados como se pode observar na figura 45.

SEARCH	Leaf type - triangular / Heartwood color - Blue / Outer bark - Flat / Root Type - submersed
RESULTS	
	Macaca-poranga
	Ypê
	Pau-rosa
	Angelim

Figura 45. Generalização de resultados

Por sua vez, caso o usuário informe um conjunto de características conflitantes, o excesso de especificação provoca a ausência de resultados (Figura 46).

SEARCH	Phytogeographic Domain - Amazon / Type of secretory element - resinous Heartwood color after photo oxidation - greyish / Stem base - dilated
RESULTS	
Not found	

Figura 46. Especificação de resultados

Desta forma, observa-se a necessidade de uma maior precisão quanto aos resultados apresentados previamente pelo sistema especialista para identificação botânica. A influência da experiência dos usuários deve ser minimizada para propor melhorias no processo de tomada de decisão.

A Construção do Vetor Semântico

O sistema especialista de suporte à tomada de decisão baseado no referencial semântico Onto-AmazonTimber atua como ferramenta no auxílio ao processo de identificação botânica. Por vezes as indicações de espécies são numerosas, entretanto podemos afirmar a contribuição no sentido de estreitar as possibilidades de espécies.

A generalização dos resultados implica na escolha do usuário dentre as espécies apresentadas, o que aumenta a margem de erro e imprecisão da identificação botânica. Em certas situações, duas espécies apresentam alta semelhança diferenciando-se por poucas características observáveis, como, por exemplo, o caso da espécie “Macaca-poranga” - *Aniba parviflora* e “Pau-rosa” - *Aniba rosaeodora* Ducke. Tal semelhança foi evidenciada pela comunidade científica botânica com estudos para caracterização e diferenciação entre as espécies.

O excesso de especificação dos resultados ocorre em circunstâncias em que o conjunto de características botânicas dispostas pelo usuário não faz referência especificamente a nenhuma espécie. Os motivos que ocasionam a ausência de resultados prendem-se à indicação de características erradas de uma espécie ou quando as características informadas não estão incluídas no referencial semântico.

Neste sentido, para melhorar a acurácia dos resultados apresentados, o presente trabalho propõe a construção de um vetor semântico para cada espécie botânica, com inclusão de pesos que evidenciam o grau de relevância de cada característica para sua identificação. É importante referir que o referido grau de relevância das características inferidas pelos pesos

foi atribuído por especialistas do cenário botânico e por consultas a literatura científica relevante.

Desta forma, o vetor semântico é representado por uma matriz de duas colunas como se pode observar na Figura 47, onde a primeira contém termos da ontologia Onto-AmazonTimber que dispõem as características desta espécie e a segunda coluna contém a qualificação deste termo, ou seja, o peso por porcentagem que representa o quão relevante esta característica é para a identificação desta espécie.

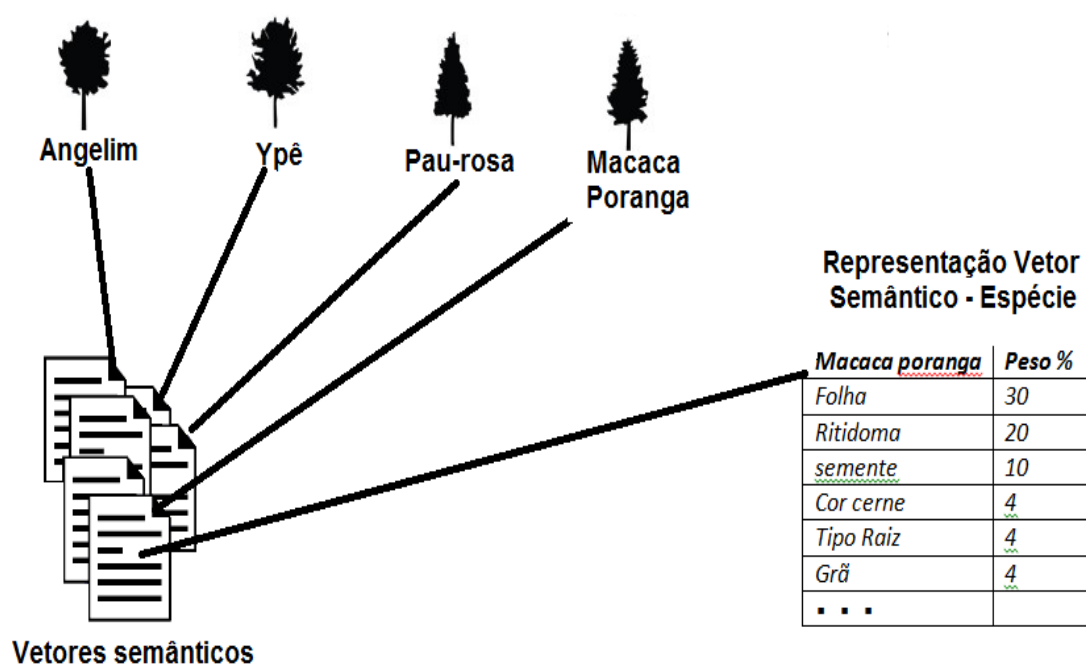


Figura 47. Representação do Vetor Semântico

5.2.2.2 MÁQUINA DE INFERÊNCIA

O mecanismo de inferência do sistema especialista para identificação botânica usa a Onto-AmazonTimber. Além disto, foram utilizados outros métodos de inferências estatísticas como o modelo estatístico da Distância Euclidiana aplicado a modelos vetoriais. Neste trabalho utiliza-se a distância euclidiana como um método para calcular o grau de similaridade entre os vetores semânticos. A distância euclidiana é definida como a distância entre dois vetores em um espaço vetorial Euclidiano.

Grau de Similaridade medido por Distância Euclidiana

Durante o processo de identificação botânica, o usuário assume a tarefa de informar características botânicas identificadas na espécie. Estas características compõem um vetor semântico com pesos equalizados de forma equivalente. Por sua vez, este vetor semântico

dispõe de uma representação gráfica na qual cada característica situa-se como eixo do produto cartesiano e o ponto de convergência entre os eixos representa esta espécie, como se pode observar na Figura 48.

Neste sentido, cada espécie observada no resultado dispõe de um vetor semântico, que contém as mesmas características indicadas pelo usuário no momento da pesquisa, que por sua vez, produzem um produto cartesiano que represente este vetor semântico. Desta forma calcula-se o nível de proximidade (similaridade) dos produtos cartesianos de cada espécie com o produto cartesiano do vetor semântico da pesquisa. Assim, a que apresentar maior similaridade corresponde à espécie indicada.

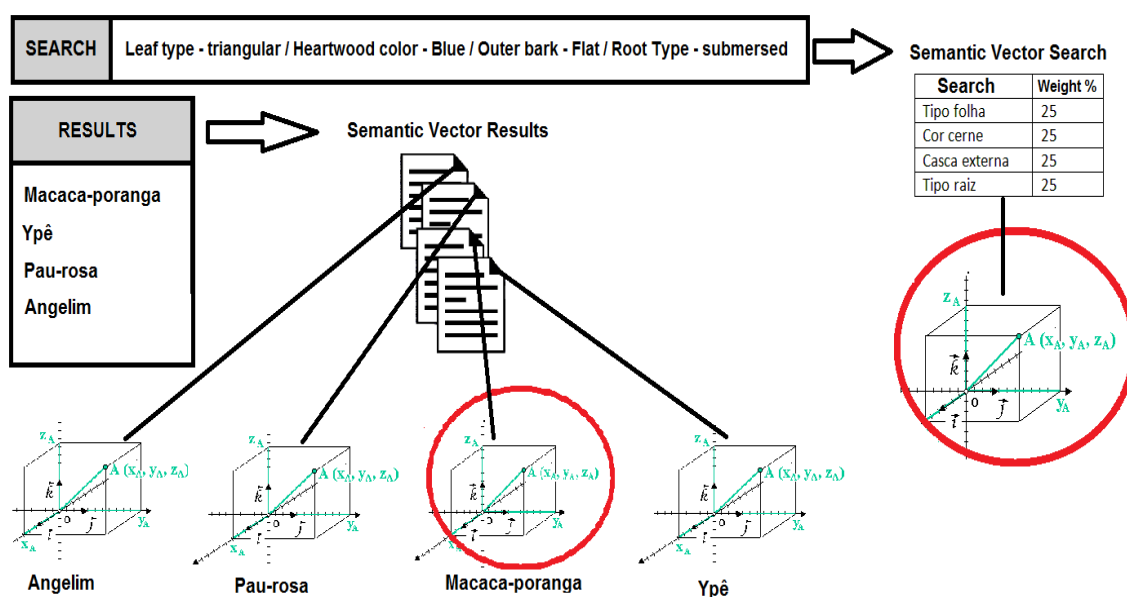


Figura 48. Exemplo de Vetor Semântico

5.2.2.4 ARQUITETURA DO SISTEMA ESPECIALISTA PARA IDENTIFICAÇÃO BOTÂNICA

Ao analisar a arquitetura (Figura 49), observa-se a presença de duas camadas que apresentam arquiteturas particulares: sistemas especialistas e ontologia. Entretanto possuem um ponto de convergência, parte de seus componentes situam-se no conjunto de interseção.

O sistema especialista é composto pela interface, base de conhecimento e máquina de inferência. A interface situa-se como área de interação com o usuário que não necessariamente precisa ser um especialista; a base de conhecimento contém o conhecimento abstraído do especialista, e por fim a máquina de inferência que permite depreender o contexto semântico e deduzir soluções. Estes componentes do sistema especialista detêm um fluxo de interação que permite sua constante evolução.

Por conseguinte, o sistema especialista possui uma base de conhecimento representada pela ontologia Onto-AmazonTimber composta por axiomas, relações semânticas, propriedades, conceitos e indivíduos, que formaliza a representação do conhecimento assim como as regras e contexto do domínio do conhecimento.

O conhecimento angariado representado por relações semânticas compostas por características e imagens botânicas permitem o diagnóstico e auxílio no processo cognitivo de tomada de decisão.

Neste sentido, a ontologia Onto-AmazonTimber situa-se como alicerce para o exercício da função do sistema especialista, promovendo:

- Melhor gerenciamento e formalização do conhecimento do especialista.
- Maior escalabilidade para expandir outros serviços no domínio de conhecimento.
- Maior facilidade para integração com outras tecnologias.

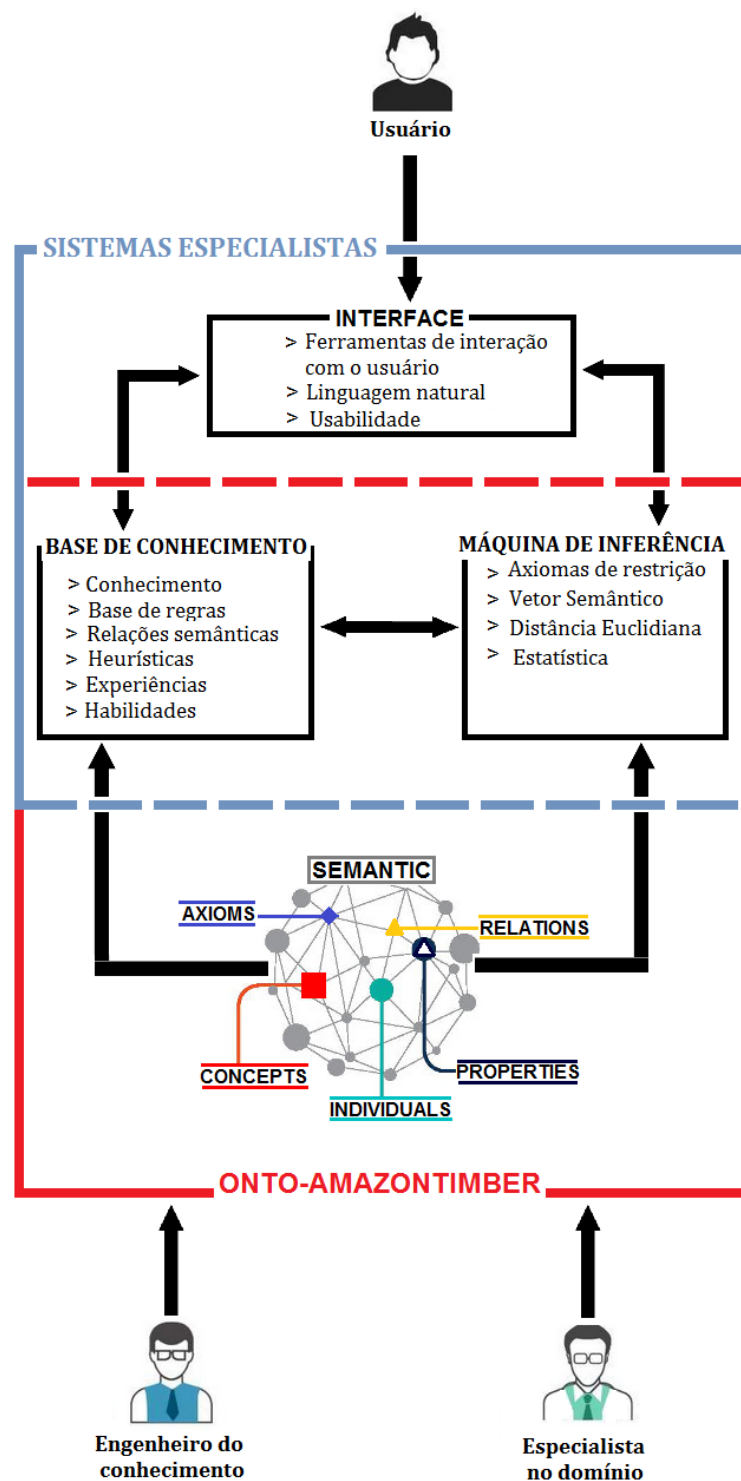


Figura 49. Arquitetura do SE

5.2.3 Sistema Especialista para classificação de imagem de madeira

As características anatômicas da madeira (e.g. raios, placas de perfuração, parênquimas, vasos e demais estruturas da mecânica da madeira) apresentam padrões específicos para cada espécie. Estes padrões podem ser observados em uma imagem

macroscópica da madeira, obtida por um simples dispositivo de maximização de imagem acoplado a um dispositivo móvel.

Neste contexto, o sistema especialista para classificação de imagens de madeira é utilizado como instrumento de validação no Cenário Imagem Madeira que objetiva identificar espécies botânicas por reconhecimento de padrões da imagem da madeira.

Angariou-se nesta pesquisa um banco de imagens macroscópicas de madeiras coletado no laboratório de madeira da Universidade Federal do Oeste do Pará (UFOPA). Tais imagens, juntamente com características de espécies e do ambiente em que estão inseridas, são parte da Onto-AmazonTimber.

Recursos de processamento de imagem são aplicados para abstrair características visuais da madeira, como cor (pigmentação), geometria e frequência dos raios (parênquimas), vasos, traqueóides, xilemas, entre outros conjuntos de primitivas. Por conseguinte, tais características formam padrões que serão mapeados por um algoritmo de classificação, no intuito de indicar a espécie que melhor adequa às características apresentadas.

5.2.3 1 ETAPA 1: ROTULAR IMAGENS (PRÉ-PROCESSAMENTO DE IMAGENS BOTÂNICAS E EXTRAÇÃO DAS CARACTERÍSTICAS DE HARALICK)

Nesta etapa são rotuladas as imagens por espécies florestais e analisa-se o grau de diferença entre as espécies. Para tal é necessário definir as propriedades capazes de caracterizar o conjunto de espécies florestais. Desta forma, utilizaram-se as características de Haralick para extrair propriedades das imagens. Após caracterizar cada imagem é necessário verificar se estas propriedades formam padrões por espécie. Deste modo aplicou-se a função de clusterização k-medoids, que objetiva formar grupos baseados nos dados gerados pelas características de Haralick, sendo que o resultado ótimo ocorre quando o agrupamento equivale à quantidade de espécies indicadas.

Extração das características

As Características de Haralick representam propriedades da textura de uma imagem através de um conjunto de medidas estatísticas. São geralmente aplicadas em imagens que apresentam uma textura assimétrica. Estas medidas estatísticas compõem um vetor de características de tamanho n , capaz de individualizar as imagens trabalhadas.

O vetor de características é estimado com base na matriz de coocorrência – matrizes em que o número de vezes que ocorreu a transição do nível de cinza z_i para z_j em uma imagem é calculada a partir de diferentes ângulos e distâncias, como indicado na figura 50. Para melhor compreensão das formulas, adotaremos a notação para a matriz de coocorrência

normalizada: $p(i,j,d,\theta)$: representa o cálculo de quantas ocorrências de passagens do pixel i para o pixel j , que está a uma distância d a uma angulação θ , esta contagem é realizada para todos os elementos da matriz. $p(i,j)$: representa o elemento (i,j) normalizado. p_x : soma dos valores da i -ésima linha. p_y : soma dos valores da j -ésima coluna. N_g : número de níveis de cinza que compõem a imagem. $p(x-y)(k)$: representa a $\sum_i \sum_j p(i,j)$, se $|i-j|=k$.

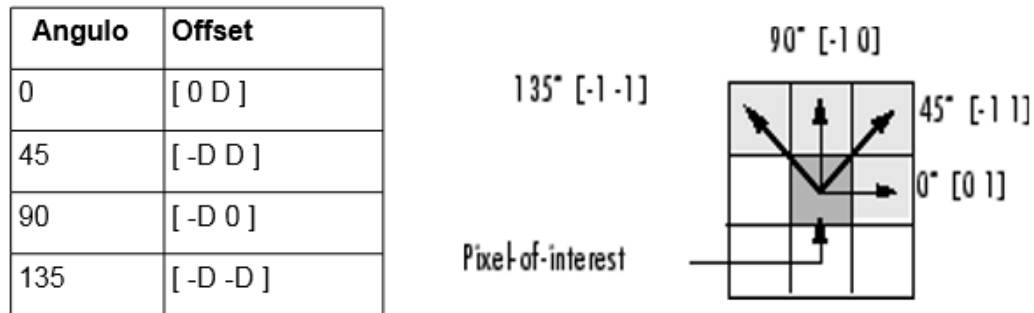


Figura 50. Matriz de coocorrência

Agrupamento - Resultados

A análise de agrupamentos é uma técnica de análise multivariada que tem por objetivo a construção de grupos (ou *clusters*) a partir de uma base de dados composta por n objetos (registros) com f atributos. A formação dos grupos ocorre pelo alto grau de homogeneidade entre os integrantes do grupo e baixo grau de homogeneidade entre integrantes de grupos diferentes. O grau de homogeneidade é aplicado por um método de grupamento que pode ser classificado por não hierárquicos e hierárquicos. Em particular, propõe-se neste trabalho o método de agrupamento não hierárquico para a resolução do problema dos k -medoids, como função de avaliação de homogeneidade utiliza-se a distância euclidiana na alocação dos objetos. Como resultado, podem-se extrair cerca de 14 características mostradas na Tabela 11.

Tabela 11. Características de Haralick

Tipos	Energia	Entropia	IDM	Inercia	Soma das Medias	Soma das Variâncias	Soma das Entropias	Diferença Medias	Diferença Variâncias	Diferença Entropias	Correlação	1 Info de Correlação	2 Info de Correlação
01	0.57641	1.1544	0.89886	0.20449	0.46263	0.64594	1.0777	0.076556	0.070011	0.3825	0.79391	-0.53898	0.75083
02	0.23322	2.5014	0.82121	0.42284	2.2129	2.1872	2.2847	0.20914	0.16314	0.72791	0.88909	-0.48262	0.88476
03	0.47653	1.4101	0.89431	0.2168	0.75453	1.0649	1.3444	0.065182	0.060473	0.34121	0.91826	-0.67365	0.86971
04	0.27286	2.229,00	0.85053	0.34195	1.6361	2.5679	2.0776	0.14488	0.12325	0.58839	0.9346	-0.59976	0.9193
05	0.28077	2.3114	0.88337	0.26059	1.7101	3.484,00	2.1795	0.13022	0.11142	0.54501	0.95348	-0.65201	0.94237
06	0.25512	2.2512	0.82635	0.38992	3.6131	2.2111	2.0795	0.16981	0.13736	0.63767	0.90893	-0.5399	0.89306
07	0.65373	1.2668	0.88929	0.24609	2.1279	0.71847	1.1668	0.097369	0.087757	0.45575	0.85832	-0.46755	0.72661
08	0.19755	2.6772	0.82144	0.42306	2.4187	3.4955	2.4791	0.19392	0.1521	0.68871	0.93203	-0.56954	0.93351
09	0.20235	2.6697	0.84613	0.36655	2.8055	4.0645	2.4847	0.17748	0.14486	0.66471	0.95277	-0.60421	0.94602
10	0.72825	0.9402	0.90194	0.20196	3.8008	0.41444	0.86681	0.072735	0.067125	0.37181	0.77221	-0.43092	0.62814
11	0.70865	0.93715	0.93088	0.1387	1.7794	0.36866	0.85784	0.079284	0.072841	0.39804	0.76461	-0.38303	0.59587
12	0.67705	0.94279	0.91987	0.16026	0.30191	0.44047	0.87055	0.072237	0.066618	0.36977	0.76728	-0.46012	0.65068
13	0.23636	2.3864	0.83706	0.37741	2.177,00	2.6489	2.2138	0.17026	0.13789	0.63985	0.9183	-0.5638	0.91424
14	0.35359	1.8733	0.90569	0.1972	1.0571	1.405,00	1.7383	0.13474	0.11439	0.55617	0.88698	-0.5321	0.85459

Os índices da variável tipo são apresentados na Tabela 12.

Tabela 12. Variável tipo da tabela 11

TIPO	NOME POPULAR	NOME CIENTÍFICO
01	Tatajuba	Bagassa guianensis
02	Cedrorana	Cedrelinga cateniformis
03	Cumaru	Dipteryx ferrea
04	Cupiúba	Goupia glabra
05	Ipê	Handroanthus sp
06	Angelim	Hymenolobium petraeum
07	Andiroba	Carapa guianensis
08	Pau-roxo	Peltogeny spp.
09	Fava	Vataieropsis sp.
10	Caju-açu	Anacardium giganteum
11	Tauarí	Couratari spp.
12	Freijó	Cordia goeldiana Huber
13	Anani	Symphonia globulifera
14	Maçaranduba	Manilkara spp.

Contudo nem todas as características são significativas para a formação dos grupamentos, contribuindo de certa forma para inconsistência dos grupos. Por exemplo, todos os valores para correlação são próximos de um, indicando que os pixels das imagens apresentam alto grau de dependência linear. Desta forma, por não apresentarem grandes variações entre si, é provável que o tipo 01 e tipo 10 sejam atribuídos ao mesmo grupamento, mesmo tratando-se de espécies diferentes. O mesmo acontece com os valores referentes à variável IDM (*Inverse Difference Moment*).

Outro problema ocorre na Soma das Médias, no qual imagens da mesma espécie apresentam valores discrepantes, ocasionando a alocação para grupos diferentes. Neste contexto, optou-se por variáveis em que a divisão dos grupamentos fosse mais evidente (Tabela 13). Foram tratadas apenas 10 imagens por espécie para que um grande número de classes não distorcesse os resultados.

Tabela 13. Agrupamento com características selecionadas

Clusters(k-medoids)	Tipos	Energia	Entropia	Soma das Variâncias	Diferença Entropias	Info de Correlação 1	Info de Correlação 2
7	01	0.57641	1.1544	0.64594	0.3825	-0.53898	0.75083
9	02	0.23322	2.5014	2.1872	0.72791	-0.48262	0.88476
2	03	0.47653	1.4101	1.0649	0.34121	-0.67365	0.86971
8	04	0.27286	2.229	2.5679	0.58839	-0.59976	0.9193
4	05	0.28077	2.3114	3.484	0.54501	-0.65201	0.94237
1	06	0.25512	2.2512	2.2111	0.63767	-0.5399	0.89306
7	07	0.65373	1.2668	0.71847	0.45575	-0.46755	0.72661
5	08	0.19755	2.6772	3.4955	0.68871	-0.56954	0.93351
3	09	0.20235	2.6697	4.0645	0.66471	-0.60421	0.94602
6	10	0.72825	0.9402	0.41444	0.37181	-0.43092	0.62814

Como se pode perceber, o tipo 10 (caju açúcar) é a imagem mais homogênea do grupo. Por serem medidas inversamente proporcionais, o tipo 10 também é a imagem com menor grau de entropia, por ser uma imagem com pouco ruído. Enquanto o tipo 13 (anani) é a imagem menos homogênea, logo a com maior valor de entropia. A imagem mais homogênea recebeu o menor valor de soma das variâncias, enquanto a segunda imagem menos homogênea (09 – 0.202335) recebeu o maior valor para soma das Variâncias. A diferença das entropias se comporta de maneira inversa à entropia, atribuindo valores mais baixos a imagens mais Entrópicas.

5.2.3.2 ETAPA 2: CLASSIFICAÇÃO DAS IMAGENS- K-MEANS X REDES NEURAIS ARTIFICIAIS

A Etapa 2 consiste em identificar espécies florestais através de métodos de agrupamento e classificação de dados. O comparativo entre os métodos permite a análise da eficácia dos resultados. O conjunto de dados consiste em 19 imagens, sendo 8 (oito) imagens referentes à espécie Maçaranduba e 11 (onze) imagens referentes à espécie Tauari. Estas imagens foram postas em tons de cinza e em seguida foram extraídas as características de Haralick, dentre as quais, a característica 'Soma das Médias' foi considerada a que melhor descreve as imagens do conjunto de dados. Após a extração da característica 'Soma das Médias', foram aplicados os método agrupamento (K-Means) e método de classificação (RNA).

Extração de Características de Haralick

As características podem ser: Energia, Entropia, Correlação, entre outras. Contudo, nem todas as características são significativas em um processo de agrupamento ou classificação. Por exemplo, para a característica Energia (Figura 51), os testes mostraram que se há duas imagens com aspecto caótico ou com aspecto homogêneo, tendem a apresentar valores de Energia próximos, no entanto podem pertencer a espécies diferentes.

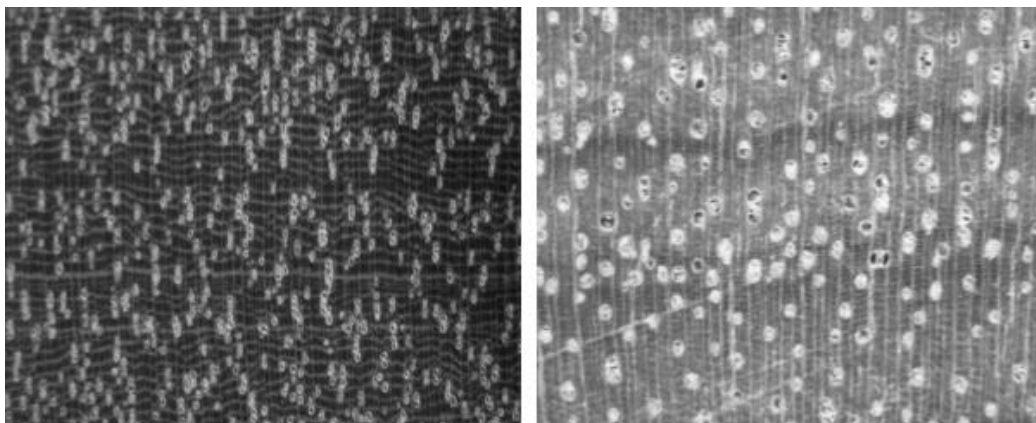


Figura 51. Respectivamente: Maçaranduba(M01):0.35515; Tauari (T06):0.3506

Contudo uma característica, dentre as analisadas, mostrou variações significativas as quais podemos associar a um grau de diferenciação entre espécies. Detaca-se "Soma das Médias", uma medida estatística dada pela função:

$$\text{Soma das Médias} = \sum_{i=2}^{2N_i} ip_{x+y}(i)$$

Esta característica foi extraída de um grupo de 19 imagens, sendo 8 de Maçaranduba e 11 de Tauari. Os respectivos valores podem ser observados na Tabela 14, onde D1_0 representa um relação de distância 1 pixel em grau 0, enquanto D1_45 representa uma relação de distância 1 pixel em grau 45.

Tabela 14. A característica Soma das Médias para o conjunto de Imagens.

Imagem	Índice	D1_0	D1_45
M01	1	1,312900	1,312700
M02	2	1,225800	1,226000
M03	3	1,881800	1,881700
M04	4	1,818900	1,819100
M05	5	0,414830	0,414830
M06	6	0,388550	0,388650
M07	7	0,534510	0,534450
M08	8	0,444700	0,444720
T01	9	3,316700	3,316700
T02	10	3,118100	3,118200
T03	11	3,184800	3,185000
T04	12	2,153100	2,153000
T05	13	2,137200	2,137200
T06	14	3,455000	3,455300
T07	15	3,160400	3,160400
T08	16	3,744200	3,744300
T09	17	3,449500	3,449600
T10	18	3,011200	3,011400
T11	19	3,079000	3,079000

Agrupamento por K-means

O Algoritmo de K-Means é um processo iterativo, que analisa a matriz de dados e calcula a distância entre o centróide calculado e a variável, geralmente por Distância Euclidiana e, assim, os dados são agrupados em torno do centróide do qual apresentam menor distância. À medida que o algoritmo itera, o centróide é recalculado e os dados são reagrupados. No decorrer das iterações a distância entre o centróides do cluster e os dados diminuem, até o ponto em que os agrupamentos parem de sofrer variações ao se atingir o critério de parada.

Observam-se várias limitações no algoritmo de K-Means, principalmente quando os clusters têm tamanho, densidade e formatos diferentes ou quando o conjunto de dados contém outliers, ou seja, é muito sensível a ruídos. Foram executadas várias vezes a função kmeans, para diferentes combinações do conjunto de dados Soma das Médias.

Para a primeira execução, foi utilizado, como parâmetro, o valor D1_0 (distância 1 na horizontal). Ao executar o K-Means, a função apresenta primeiramente uma atribuição de clusters inadequada para a espécie Tauari. Ao ser reexecutada, a função retorna valores corretos, mas a incerteza da resposta do algoritmo não é adequada para o problema. O índice gerado pode ser observado na Tabela 15.

Tabela 15. Agrupamento com K-means para Tauari e Maçaranduba (primeira execução)

Primeira Execução						
Imagem	Índice	D1_0	Distância de C1	Distância de C2	Clusters	Centroides
M01	2	1,312900	3,8690015783	0,0066701522	C1	3,2798777778
M02	2	1,225800	4,2192355172	2,9474041e-05	C2	1,231229
M03	2	1,881800	1,9546214727	0,423242626		
M04	2	1,818900	2,1344560672	0,3453572042	Clusters	Sumd
M05	2	0,414830	8,2084987689	0,6665073272	C1	0,4384168956
M06	2	0,388550	8,3597763185	0,710107897	C2	4,9265895071
M07	2	0,534510	7,5370442353	0,485417365		
M08	2	0,444700	8,0382330316	0,6186278678		
T01	1	3,316700	0,001355876	4,3491892918		
T02	1	3,118100	0,0261720494	3,5602821706		
T03	1	3,184800	0,0090397838	3,816439652		
T04	2	2,153100	1,2696281605	0,8498461406		
T05	2	2,137200	1,3057125038	0,8207834528		
T06	1	3,455000	0,0306677927	4,9451574604		
T07	1	3,160400	0,0142749394	3,7217007472		
T08	1	3,744200	0,215595126	6,3150232468		
T09	1	3,449500	0,0287716983	4,9207262294		
T10	1	3,011200	0,0721877483	3,1682967608		
T11	1	3,079000	0,0403518816	3,4142576684		

Percebe-se que, durante a primeira execução, o grupo da Maçaranduba foi agrupado corretamente, enquanto apenas 2 espécies de Tauarí foram agrupadas junto às Maçarandubas, sendo que C1 e C2 são os centróides calculados como base. A tabela 16 em que as espécies T04 e T05 estão mais próximas do centróide do grupo 2 (a uma distância de 0,8498461406u e 0,8207834528u) que do centróide do grupo 1 (a uma distância de 1,2696281605 e 1,305725038, respectivamente).

Contudo, observando-se a segunda execução, percebe-se que foram produzidos bons resultados de atribuição dos grupos. Observa-se, desta forma, o agrupamento correto de todos os valores da variável D1_0. Pode-se perceber que há um valor diferente calculado para os centróides, as imagens T04 e T05 antes agrupadas em um grupo errado, agora estão mais próximas do centróide do grupo 1 e mais distantes do centróide do grupo 2.

Tabela 16. Agrupamento com K-means para Tauarí e Maçaranduba (segunda execução)

Segunda Execução						
Imagem	Indice	D1_0	Distância de C1	Distância de C2	Clusters	Centroides
M01	2	1,312900	3,0999364404	0,0961937979	C1	3,0735636364
M02	2	1,225800	3,4142304559	0,0497518601	C2	1,00274875
M03	2	1,881800	1,420300565	0,7727311001		
M04	2	1,818900	1,5741808404	0,6661028629	Clusters	Sumd
M05	2	0,414830	7,0688645491	0,3456484566	C1	2,5455367855
M06	2	0,388550	7,2092982275	0,3772401045	C2	2,8383341165
M07	2	0,534510	6,4467933683	0,219247527		
M08	2	0,444700	6,9109240186	0,3114184074		
T01	1	3,316700	0,0591152913	5,3543703874		
T02	1	3,118100	0,0019834877	4,4747109109		
T03	1	3,184800	0,0123735286	4,7613476576		
T04	1	2,153100	0,8472533059	1,3233079984		
T05	1	2,137200	0,8767768595	1,2869796386		
T06	1	3,455000	0,1454936995	6,0135361931		
T07	1	3,160400	0,007540554	4,6554589166		
T08	1	3,744200	0,4497531322	7,5155549561		
T09	1	3,449500	0,1413281495	5,9865916794		
T10	1	3,011200	0,0038892231	4,0338764236		
T11	1	3,079000	2,9554049e-05	4,3108192531		

Quando se executa o método k-means (Tabela 17) com duas variáveis, resultados incorretos são produzidos, tanto na primeira quanto na segunda execução. Através disso, podemos observar que não há uma melhora significativa dos resultados quando se acrescenta uma nova variável.

Tabela 17. Agrupamento com K-means para Tauari e Maçaranduba (com duas variáveis)

Primeira Execução								
Imagem	Índice	D1_0	D1_45	Distância de C1	Distância de C2	Clusters	C_D1_0	C_D1_45
M01	2	1,312900	1,312700	7,7392271506	0,0133066985	C1	3,2798777778	3,27998889
M02	2	1,225800	1,226000	8,4381058728	5,6879266e-05	C2	1,231229	1,231235
M03	2	1,881800	1,881700	3,9098332895	0,8463473423			
M04	2	1,818900	1,819100	4,2686524128	0,6909424625	Clusters	Sumd	
M05	2	0,414830	0,414830	16,4176342275	1,3330244512	C1	0,8768878844	
M06	2	0,388550	0,388650	16,719616889	1,4200573793	C2	9,8529485171	
M07	2	0,534510	0,534450	15,0750280257	0,9709267012			
M08	2	0,444700	0,444720	16,0769827039	1,2372337131			
T01	1	3,316700	3,316700	0,0027035817	8,6983535581			
T02	1	3,118100	3,118200	0,052347694	7,1209190819			
T03	1	3,184800	3,185000	0,0180626728	7,6336373273			
T04	2	2,153100	2,153000	2,5397321162	1,6994968559			
T05	2	2,137200	2,137200	2,6116789484	1,6415560341			
T06	1	3,455000	3,455300	0,0614017784	9,8916225847			
T07	1	3,160400	3,160400	0,0285764417	7,4433783445			
T08	1	3,744200	3,744300	0,431179934	12,6305189411			
T09	1	3,449500	3,449600	0,0575396273	9,8418695027			
T10	1	3,011200	3,011400	0,1443277395	6,3372841881			
T11	1	3,079000	3,079000	0,0807484151	6,8284931637			

Classificação por Redes Neurais Artificiais

Para a execução da classificação das espécie Maçaranduba e Tauari, utilizou-se RNA, que apresenta a seguinte configuração: cinco neurônios na camada escondida (*rede = newff(Limites, [5 1] , {'tansig','purelin'},'trainlm')*); 1000 épocas para treinamento com exibição a cada 200 épocas (*rede.trainParam.epochs = 1000; rede.trainParam.show = 200*); taxa de erro de 0,000000001 (*rede1.trainParam.goal = 1e-8*), onde o algoritmo de treinamento utilizado foi de Levenberg-Marquardt (LEVENBERG, 1944)(MARQUARDT, 1963).

A classificar por RNA utiliza-se como conjunto de treinamento as imagens: [M04 M08 M06 M11], sendo que as demais imagens são utilizadas como conjunto de teste. O padrão de entrada indica que deve ser atribuído 0 (zero) quando a imagem for Maçaranduba ou 1 (um) quando a imagem for Tauari (Tabela 18)

Tabela 18. Classificação por RNA para Tauarí e Maçaranduba

Rede Neural para Um Parametro					
Imagem	Índice	D1_0	C. Treinamento	C. Teste	Classificação
M01	1	1,312900		X	-0,1902614135
M02	2	1,225800		X	-0,1784318351
M03	3	1,881800		X	0,0806471002
M04	4	1,818900	X		-
M05	5	0,414830		X	0,0015645434
M06	6	0,388550		X	0,0029406508
M07	7	0,534510		X	-0,0052722826
M08	8	0,444700	X		-
T01	9	3,316700		X	1,0198199483
T02	10	3,118100		X	1,0083218647
T03	11	3,184800		X	1,0173580180
T04	12	2,153100		X	0,4646281155
T05	13	2,137200		X	0,4469891771
T06	14	3,455000	X		-
T07	15	3,160400		X	1,0147382078
T08	16	3,744200		X	0,8033127476
T09	17	3,449500		X	1,0013537280
T10	18	3,011200		X	0,9790973143
T11	19	3,079000	X		-

No agrupamento com a utilização do método k-means, ocorrem falhas para T04 e T05 quanto ao valor apresentado para a classificação destas imagens. Possivelmente, com um conjunto maior de treinamento, esse erro não ocorreria. Ao se aplicar a rede neural para dois parâmetros de classificação percebe-se que as imagens T04 e T05 (Tabela 19) que antes apresentavam valores com tendência a 0 (zero) agora estão claramente próximos a 1, logo estão classificados corretamente. Percebe-se que diferentemente do K-Means, a Rede Neural se aprimora e se torna mais precisa.

Tabela 19. Classificação por RNA para Tauarí e Maçaranduba (com duas variáveis)

Rede Neural para Dois Parametros						
Imagem	Índice	D1_0	D1_45	C. Treinamento	C. Teste	Classificação
M01	1	1,312900	1,312700		X	-1,8850765794
M02	2	1,225800	1,226000		X	-1,9074811808
M03	3	1,881800	1,881700		X	0,2382836851
M04	4	1,818900	1,819100	X		-
M05	5	0,414830	0,414830		X	0,0293538874
M06	6	0,388550	0,388650		X	0,0514507852
M07	7	0,534510	0,534450		X	-0,1194198480
M08	8	0,444700	0,444720	X		-
T01	9	3,316700	3,316700		X	1,0002678207
T02	10	3,118100	3,118200		X	1,0001782488
T03	11	3,184800	3,185000		X	1,0002742354
T04	12	2,153100	2,153000		X	0,8104625522
T05	13	2,137200	2,137200		X	0,7930553658
T06	14	3,455000	3,455300	X		-
T07	15	3,160400	3,160400		X	1,0002543035
T08	16	3,744200	3,744300		X	0,9995394809
T09	17	3,449500	3,449600		X	1,0001025591
T10	18	3,011200	3,011400		X	0,9997553688
T11	19	3,079000	3,079000	X		-

5.2.3.3 ETAPA 3: AVALIAÇÃO DA QUALIDADE DAS AMOSTRAS DE MADEIRA

O Método

Os corpos de prova foram polidos, utilizando-se uma lixadeira de cinta portátil, nas granulometrias de 40, 80 e 120 mesh onde, em cada fase do polimento, obtiveram-se imagens, utilizando microscópio digital com acesso via USB com câmera de 2.0 megapixel, modelo *Eletronic Magnifier*, com ampliação padrão de 500x. A aquisição das imagens apresenta 600×600 pixels no espectro visível em um quadrante de amostra medindo $0,5 \times 0,5$ cm como elemento de escala para a caracterização de dimensões da madeira. Para tal a distância entre a amostra de madeira e o microscópio digital na captura da imagem mede aproximadamente 7 cm. Assim, trabalhou-se com um total de 600 imagens, sendo 60 por espécie e 20 por tratamento.

O trabalho de reconhecimento de padrões foi dividido em três etapas, nomeadamente a segmentação da imagem, a extração de características e a classificação, que trata do reconhecimento propriamente dito (Figura 52). Na segmentação da imagem, eliminam-se informações que não sejam úteis e que possam vir a dificultar o trabalho de reconhecimento. Na extração de características, foram retirados os objetos a serem reconhecidos, visando a posterior classificação.

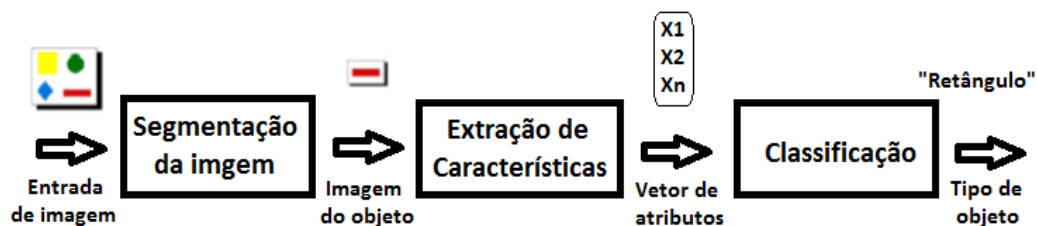


Figura 52. As três fases do reconhecimento de padrões de imagens digitais

Fonte: Adaptado de (CASTLEMAN, 1996).

Para a segmentação, optou-se por rotina do processamento de imagem com a segmentação por texturização baseado na matriz de co-ocorrência de níveis de cinza.

Na etapa de extração de características utilizou-se, como descritor de textura, as características Haralick, onde são definidas diversas características advindas do cálculo de matrizes de co-ocorrência, que são matrizes que contam as ocorrências de níveis de cinza em uma imagem. Essas características servem como medida para a diferenciação de texturas que não seguem um determinado padrão de repetitividade, fornecendo informações relevantes para classificação das mesmas. As características angariadas no descritor de Haralick, entre outras, são: homogeneidade, probabilidade máxima, entropia, momento de diferenças ordem k , momento inverso de diferença de ordem k , variância inversa, energia, contraste, variância, correlação, entre outros descritores.

Na etapa de classificação ou reconhecimento de padrão, utilizaram-se as RNAs, também conhecidas como métodos conexionistas, por meio da técnica de *Multilayer Perceptron* (MLP), com o algoritmo de treinamento supervisionado *backpropagation*. Neste algoritmo, a aprendizagem foi realizada em duas fases. Na primeira, a *forward*, os valores de saída da rede são calculados a partir dos valores de entrada fornecidos. Na segunda, *backward*, os pesos associados a cada conexão são atualizados conforme as diferenças entre os valores de saída obtidos e os valores desejados.

Do total de 20 imagens por espécie para cada tratamento, foram utilizadas 15 para treinamento da rede e 5 para teste e validação. A arquitetura da RNA desenvolvida (figura 53) apresenta como entrada uma imagem composta por descritores de textura de Haralick, perpassando pela camada oculta com sinapses neuronais para posterior aprendizagem supervisionada com a referente espécie botânica, que neste caso varia de 0 a 9.

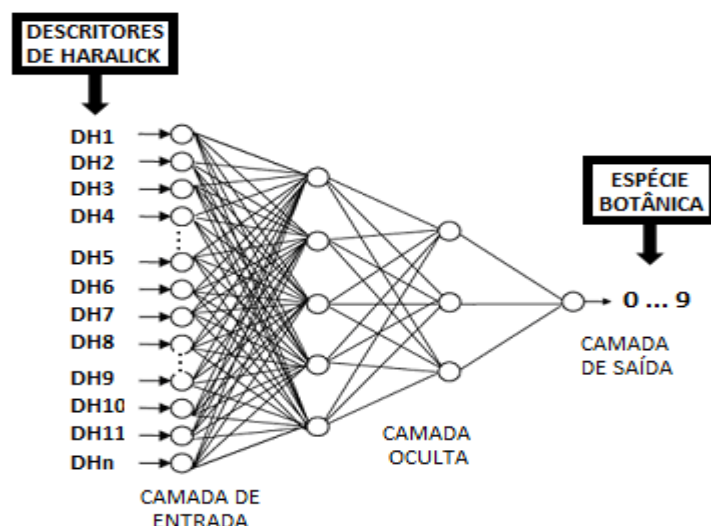


Figura 53. Arquitetura da RNA desenvolvida

O método experimental aplicado a este trabalho repete-se de forma igualitária quanto à segmentação da imagem (matriz de co-ocorrência), extração de características (descritores de Haralick) e reconhecimento de padrão (classificação por RNA) para os três cenários de teste (imagens lixadas por 40, 80 e 120).

Para análise do desempenho de treinamento da RNA, utiliza-se o gráfico de regressão linear, submetendo-o à adição de dados dispersos para enquadramento e classificação em uma linha objetivo. Isto permite avaliar o grau de reconhecimento de padrão na fase de treinamento da RNA.

Para análise do reconhecimento de padrão, atribui-se para o teste da RNA as imagens não utilizadas no treinamento que, após execução, verifica-se a saída da RNA pós-treinamento comparando-os com os resultados inferidos pelos especialistas. Desta forma constata-se a taxa de reconhecimento, ou seja, o porcentagem de imagens identificadas de forma correta pela perspectiva da RNA.

A RNA apresenta a seguinte configuração (Figura 54): dez neurônios na camada escondida (*rede = newff(Limites, [10 1] , {'tansig','purelin'},'trainlm')*); 500 épocas para treinamento com exibição a cada 50 épocas (*rede.trainParam.epochs = 500; rede.trainParam.show = 50*); taxa de erro de 0,000000001 (*rede1.trainParam.goal = 1e-8*), onde o algoritmo de treinamento utilizado foi de Levenberg-Marquardt (LEVENBERG, 1944)(MARQUARDT, 1963).

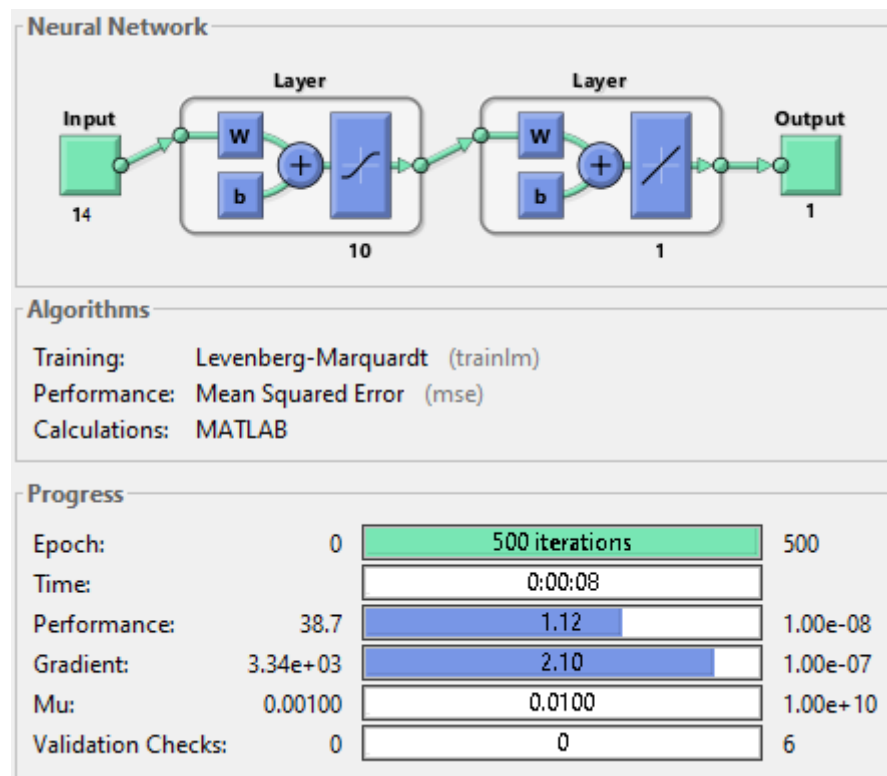


Figura 54. Configuração da RNA

Os Resultados Produzidos

Por meio da análise de desempenho de treinamento da RNA, verificou-se o enquadramento e classificação dos dados dispersos apresentados para o treinamento (Figura 55). Tal comportamento do treinamento evidencia a capacidade de reconhecimento de padrão da RNA com menor dispersão no terceiro tratamento.

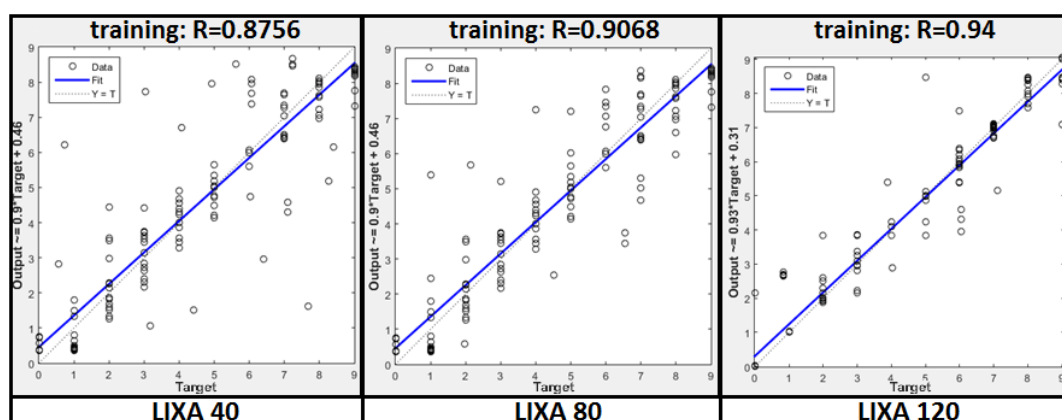


Figura 55. Gráfico de regressão linear para lixas 40, 80 e 120 respectivamente.

A etapa de treinamento da RNA situa-se como ponto crítico do desenvolvimento da aplicação, visto que nesta fase o conhecimento do especialista é transferido para o sistema de reconhecimento que, para avaliar o grau de reconhecimento das espécies florestais contidas na base de imagem, se faz necessário analisar o desempenho da RNA nesta fase de treinamento.

Quanto ao gráfico de regressão, após treinamento da RNA, observa-se (figura 55) desempenho de reconhecimento inerente ao treinamento que foi de 87%, 90% e 94% de reconhecimento em dados dispersos para os tratamentos em 40, 80 e 120 mesh respectivamente, atingindo o objetivo de similaridade com a função de regressão.

Por meio da análise dos resultados finais obtidos pela saída da RNA, verificou-se que o sistema de reconhecimento de padrão em imagens de madeira apresentado neste trabalho é capaz de identificar as espécies compostas na base de imagem com acurácia significativa, sendo estas de 55%, 62% e 65% para os tratamentos em 40, 80 e 120 mesh respectivamente (Figura 56).

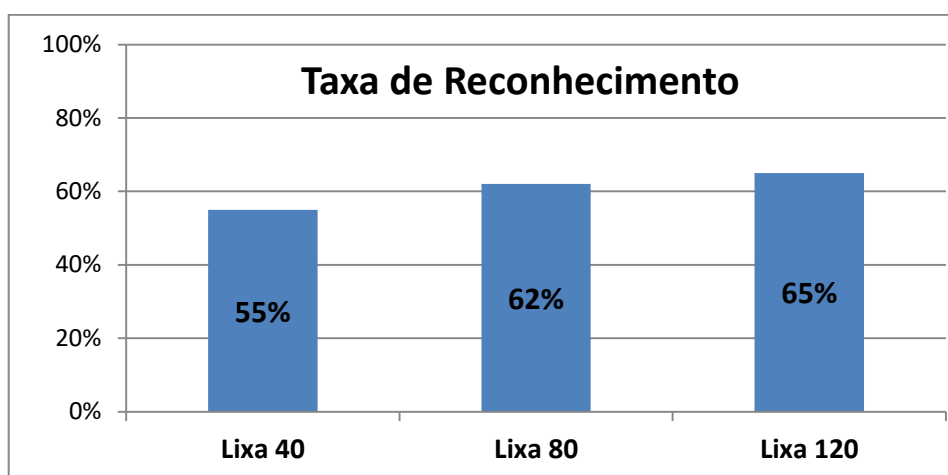


Figura 56. Taxa de reconhecimento

A aplicação de lixa 40, 80 e 120 causa um ganho gradativo quanto à retirada de características externas como ranhuras, defeitos, ondulações e marcas provenientes do corte da madeira (Figura 57), além de ganhos quanto à nitidez, luminosidade e brilho da superfície na mesma, o que repercute no desempenho da extração de características para a identificação anatômica da madeira.

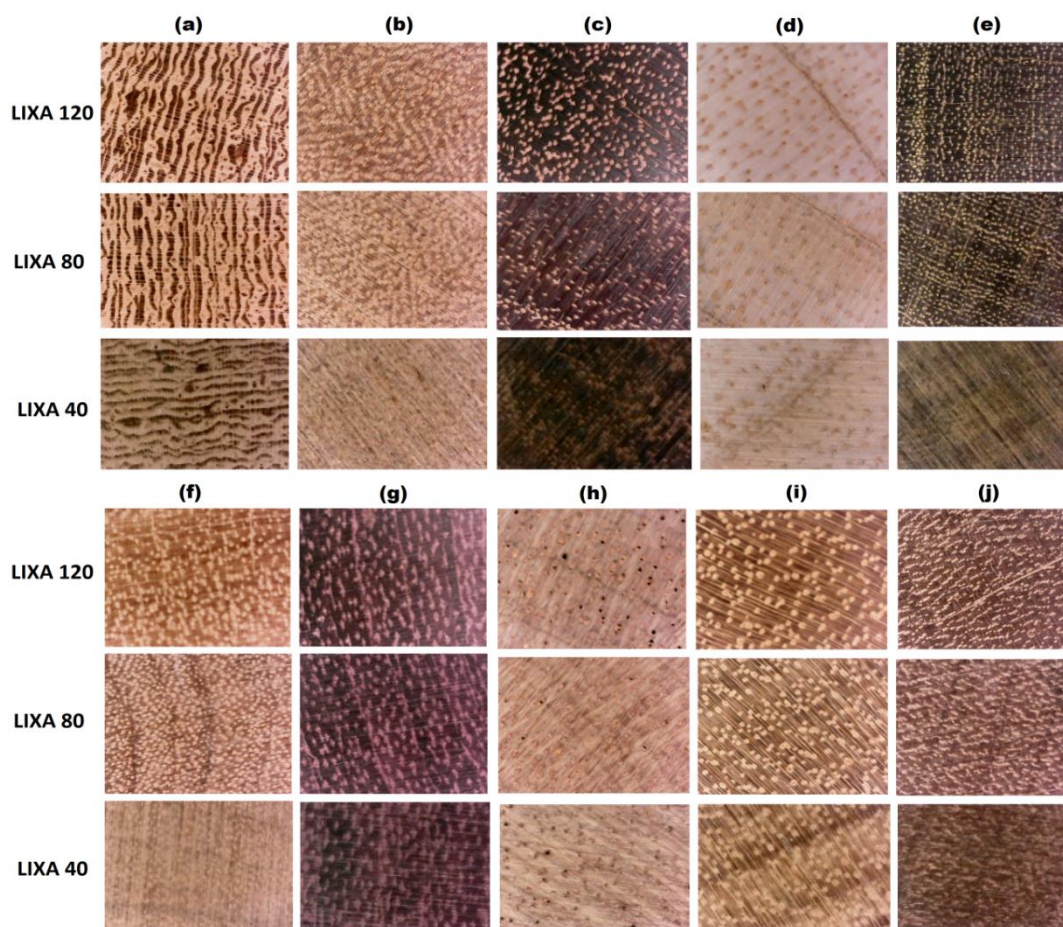


Figura 57. Imagens macroscópicas de madeiras amazônicas, dispostas por aplicação de lixa: (a) *Bagassa guianensis* (tatajuba); (b) *Carapa guianensis* (andiroba); (c) *Cedrelinga cateniformis* (cedrorana); (d) *Dipteryx ferrea* (cumaru); (e) *Goupia glabra* (cupiúba); (f) *Handroanthus* sp. (Ipê); (g) *Hymenolobium petraeum* (angelim); (h) *Manilkara* spp. (massaranduba); (i) *Peltogeny* spp. (Pau-roxo); (j) *Vataieropsis* sp. (fava).

Observa-se que nos três experimentos a taxa de reconhecimento apresenta resultados satisfatórios, comprovando a capacidade de reconhecimento das espécies florestais. No entanto, com o refinamento da superfície da madeira mediante a aplicação das lixas, a taxa de reconhecimento aumenta.

5.4 Análise dos Resultados & Validação

5.4.1 Restrições e delimitação do escopo

A validação do trabalho de tese contempla, unicamente, o referencial semântico, confirmando a execução do modelo conceitual na forma da ontologia Onto-AmazonTimber, assim como o acesso e interação com os sistemas que a utilizam.

A análise dos resultados restringe-se à quantidade de espécies formalizadas na Onto-AmazonTimber em função da delimitação do espaço que contempla a área da flona do Tapajós e das espécies devidamente catalogadas. Tal restrição aumenta em se tratando do sistema de imagens de madeira em função das amostras de madeiras disponíveis.

5.4.2 Validação Ontologia ONTO-AMAZONTIMBER

A validação do referencial semântico da pesquisa objetiva assegurar que a composição entre a estrutura semântica e conhecimentos angariados, atende aos requisitos e às especificações dos cenários de identificação botânica, e verifica se os resultados alcançados apresentem uma taxa de acurácia apropriada para o trabalho.

A Figura 58 apresenta a solução proposta neste trabalho. O referencial semântico é uma Ontologia, que tem como função formalizar o conhecimento que envolve o processo de identificação botânica, contendo conceito, indivíduos, propriedades, relações e axiomas que integram e delimitam as relações de conhecimentos na camada de suporte à tomada de decisão. Os cenários de aplicação utilizam-se de reconhecimento de padrões de características e imagens extraídos da ontologia obtidas com axiomas e recursos de processamento de imagem. Tais cenários atuam no auxílio à tomada de decisão no processo de identificação botânica, dividindo-se por contexto do manejo florestal mais especificamente no auxílio da identificação botânica na atividade de inventário florestal e no contexto da fiscalização no auxílio da identificação da espécie botânica de toras de madeira por meio de extração de imagens.

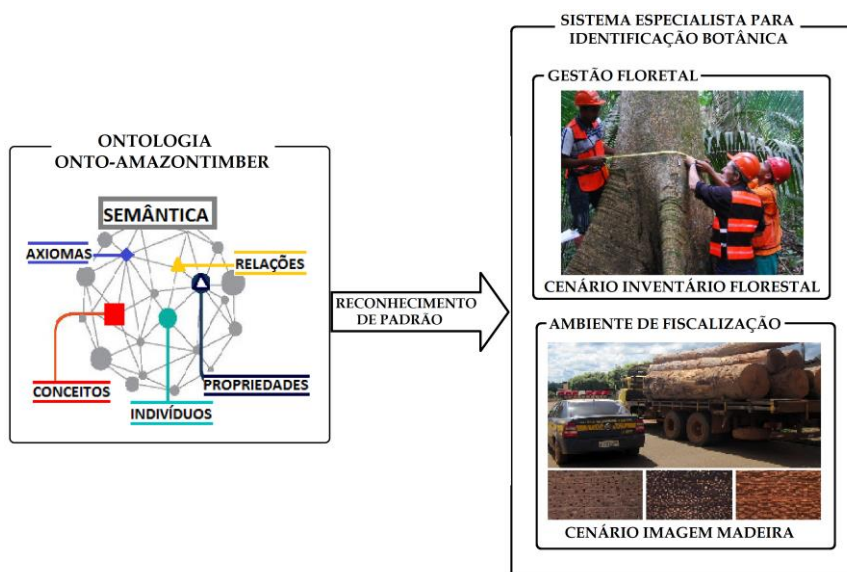


Figura 58. Validação Ontologia Onto-AmazonTimber

5.4.2 Análise dos resultados do Sistema Especialista para identificação botânica – Inventário Florestal

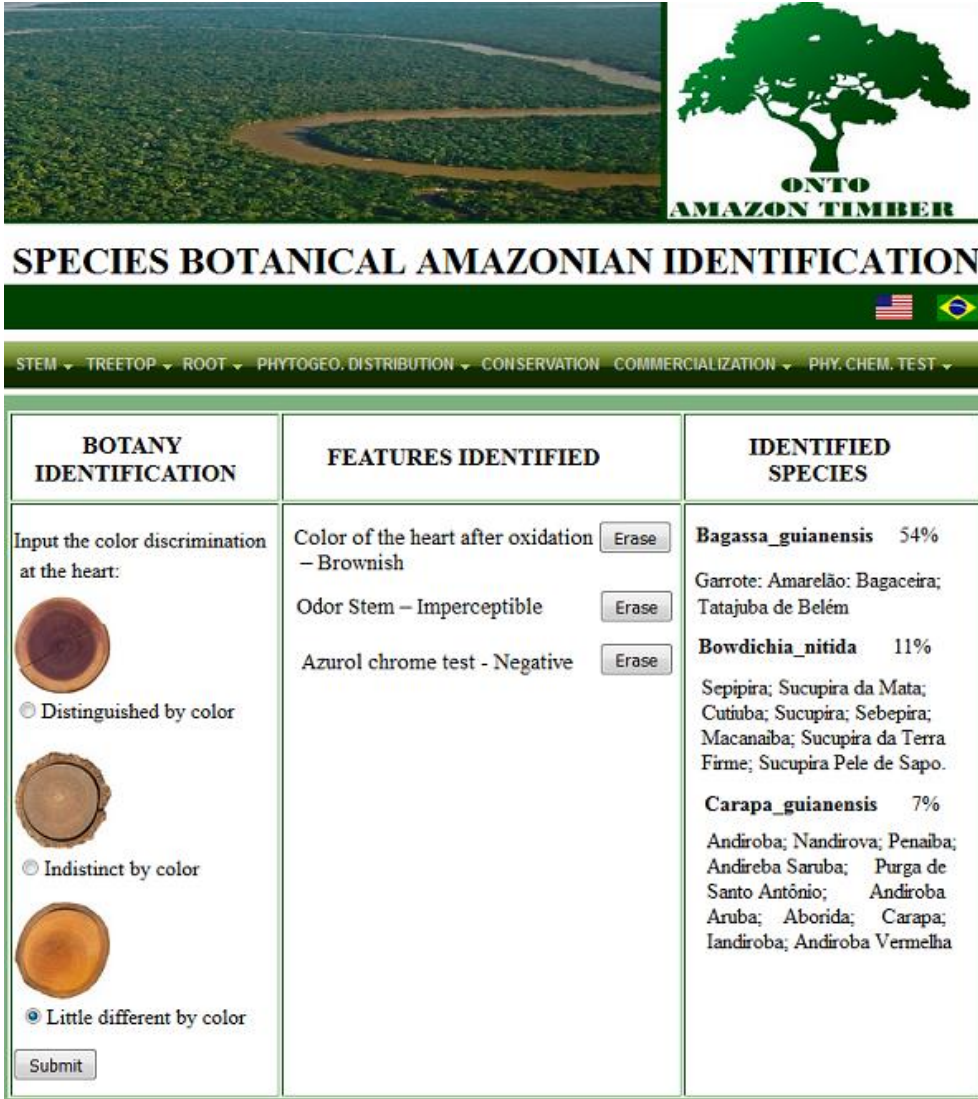
Este trabalho introduz uma abordagem de resultados baseados na utilização de vetores semânticos com auxílio da distância euclidiana, concretizados com a demonstração do grau de relevância de cada característica das espécies catalogadas, o que repercute no resultado final com a porcentagem de indicação para o diagnóstico de identificação botânica no auxílio à tomada de decisão.

Os resultados podem ser observados na Figura 59, na qual foram identificados três fatores da espécie em estudo, são estes: negatização no teste de cromazurol, odor do caule imperceptível e coloração amarronzada do cerne após oxidação. Como resultados foram identificadas três possíveis espécies: *Bagassa Guianensis*, *Bowdichia Nitida* e *Carapa Guianensis*, dispostas respectivamente quanto ao grau de similaridade e relevância com as características identificadas pelo usuário.

Observa-se no resultado que para cada nome científico estão dispostos os nomes populares que lhes fazem referência, e isto possibilita amenizar a divergência de conhecimento entre os taxonomistas com os nomes científicos e os mateiros com os nomes populares. A integração entre os conhecimentos por vias tecnológicas permite gerir e avaliar inconsistências e imprecisões provenientes deste contexto.

O diagnóstico das características da espécie botânica apresenta vários níveis de especificação, o que possibilita a identificação por usuários com vários níveis de experiências. Desta forma, quanto maior o nível de experiência do usuário maior o nível de precisão do

diagnóstico do sistema especialista. Isto posto, utilizam-se imagens no auxílio da avaliação do usuário das características como se pode observar na discriminação da cor no cerne na Figura 55. O auxílio da figura minimiza erros e inconsistências na identificação das características assim como possibilita um usuário não especialista a inferir sobre o sistema proposto.






BOTANY IDENTIFICATION	FEATURES IDENTIFIED	IDENTIFIED SPECIES
Input the color discrimination at the heart:  <input type="radio"/> Distinguished by color  <input type="radio"/> Indistinct by color  <input checked="" type="radio"/> Little different by color <input type="button" value="Submit"/>	Color of the heart after oxidation <input type="button" value="Erase"/> – Brownish Odor Stem – Imperceptible <input type="button" value="Erase"/> Azuroil chrome test - Negative <input type="button" value="Erase"/>	Bagassa guianensis 54% Garrote: Amarelão; Bagaceira; Tatajuba de Belém Bowdichia nitida 11% Sepipira; Sucupira da Mata; Cutiuba; Sucupira; Sebeipira; Macanaiba; Sucupira da Terra Firme; Sucupira Pele de Sapo. Carapa guianensis 7% Andiroba; Nandirova; Penaiba; Andireba Saruba; Purga de Santo Antônio; Andiroba Aruba; Aborida; Carapa; Iandiroba; Andiroba Vermelha

Figura 59. Interface – Sistema Especialista

5.4.1.3 ANÁLISE DOS RESULTADOS DO SISTEMA ESPECIALISTA PARA CLASSIFICAÇÃO DE IMAGENS DE MADEIRA

O processo de aquisição de imagem repercute em toda estrutura do sistema de reconhecimento de padrão. A imagem como foco principal do sistema apresenta características que variam conforme a tecnologia empregada para sua captura. Assim,

empregou-se, neste trabalho, imagens macroscópicas para reconhecimento de padrão em madeira.

A segmentação da imagem, por sua vez, é a etapa mais delicada do processamento da imagem digital, tendo em vista que o reconhecimento da imagem obtida será baseado na imagem segmentada.

Ao analisar o procedimento de aplicação de lixas nas amostras de madeiras, observa-se que, quanto maior o tratamento da superfície da madeira seja por artifícios mecânicos como aplicação de lixas ou por artifícios computacionais com aplicação de segmentação na imagem, causam ganhos significativos nos resultados de reconhecimento de padrão. No decorrer do processo de aplicação de lixa constatam-se as seguintes melhorias evidenciadas na imagem:

Com a aplicação da lixa 40, observa-se (Figura 53) a remoção de ranhuras, ondulações e defeitos provenientes do equipamento utilizado para cortar a madeira. Desta forma, a superfície da madeira passa a apresentar maior homogeneidade quanto ao relevo da zona de captura da imagem. No entanto as características anatômicas da madeira aparecem de forma discreta ou mesmo imperceptíveis mesmo em microscópio nos padrões metodológicos deste trabalho.

A aplicação da lixa 80 (figura 57) tem por objetivo limpar a superfície de visualização, retirando pequenos defeitos como linhas de ranhuras. Desta forma, a superfície da madeira passa a evidenciar com maior clareza as características da madeira, aumentando a nitidez da imagem capturada. No entanto a imagem capturada apresenta falta de luminosidade e brilho, provenientes da superfície com propriedades opaca e levemente áspera.

A lixa 120 tem como finalidade o acabamento, da sua aplicação objetiva a remoção de quaisquer defeitos que ainda perduram na superfície após a lixa 80, aumentando o grau de nitidez, aumentando consequentemente a percepção das características anatômicas da madeira. A aplicação da lixa 120 propicia o polimento da superfície o que repercute no aumento do brilho e luminosidade evidenciados na imagem obtida da madeira (Figura 57). Outras técnicas de tratamento de superfície podem ser aplicadas, a exemplo de plainagem por micrótomo ou polimento com granulometria crescente até 1200 mesh.

No entanto a aplicação das lixas utilizadas na metodologia, inclusive com o maquinário empregado, permite a possível execução em atividades práticas de fiscalização e monitoramento de cargas de madeira de maneira fácil e rápida, com o emprego de um gerador de energia, haja vista que em estradas na Amazônia muitas vezes não há acesso à mesma, três lixadeiras de cinta, lupa digital e o sistema de reconhecimento devidamente instalado no notebook ou celular.

Os resultados da taxa de reconhecimento representam o equilíbrio entre a quantidade e relevância das características abstraídas da imagem, integrado com a capacidade do método de classificação em reconhecer padrões com eficácia. Os sistemas de reconhecimento de padrões em imagens de madeira observados na literatura apresentam divergências quanto à forma de segmentação, extração de características, classificação assim como objetivos, abrangência e foco de aplicações para diferentes regiões florestais. Contudo todos os trabalhos buscam por melhores resultados de reconhecimento. Neste sentido, os trabalhos observados apresentam índices de taxa de reconhecimento com alcance em limiares de 60% à 99%. A exemplo disto, TOU (2007) com 60%, De Paula Filho e colaboradores (2014) com 65%, De Paula e Tusset (2009) com 80,9%, Xuebing (2005) com 88%, Bihui e equipe (2010) com 91,7%, Sun Lingjun e colaboradores (2011) com 93,3%, Khalid (2008) com 95%.

Em concordância com os resultados observados na literatura, este trabalho apresenta índices de taxa de reconhecimento com alcance limiar de 65%. Uma taxa de reconhecimento que atinge de forma satisfatória a tarefa de identificação botânica, no entanto precisa de melhorias para o aumento da acurácia. Para tal, são necessários outras possibilidades de acréscimos na metodologia do processo de reconhecimento de padrão, com possíveis incrementos de extração de características e outras técnicas de classificação, a citar características provenientes da cor e geometria, tais como apresentado por Sun Lingjun e equipe (2011), Bihui e colaboradores (2010) e Khalid (2008).



6. CONCLUSÕES E TRABALHOS FUTUROS

6.1 Visão Geral da Tese

As tecnologias semânticas suportam o processo de produção de conhecimento, assim como permitem a inserção de novas tecnologias no cenário de aplicação. A ontologia Onto-TimberAmazon desenvolvida nesta pesquisa apresenta um modelo conceitual embasado em um referencial semântico no domínio da botânica, no qual se integram as tecnologias de reconhecimento de padrões, representadas por axiomas e relações semânticas.

No intuito de validar o referencial semântico, foram desenvolvidos cenários de testes, nomeadamente um sistema especialista para identificação botânica e um sistema especialista para classificação de imagens de madeira. Tais cenários objetivam aperfeiçoar a tomada de decisão no processo de identificação botânica de espécies amazônicas, possibilitando, desta forma, o aumento da acurácia na identificação botânica, minimizando assim as divergências de conhecimento entre mateiros e taxonomistas.

O sistema especialista para identificação botânica (inventário florestal) apresenta os requisitos necessários para auxiliar a tomada de decisão no processo de identificação botânica das espécies comercializadas na Amazônia. Propicia a criação um cenário de inventário florestal confiável, identificando espécies florestais com maior acurácia e precisão que a tecnologia permite, portanto atenuando a margem de erro rotineira contida na prática da identificação botânica.

Por conseguinte, a estrutura do sistema especialista permite minimizar a divergência de conhecimento entre os mateiros e os taxonomistas, contribuindo para a harmonização entre os nomes científicos e os nomes populares.

Os recursos visuais apresentados por imagens das características externas das espécies botânicas, advindas da ontologia Onto-AmazonTimber propiciam uma diminuição nas incertezas das escolhas do usuário, assim como permite usuários com pouca experiência na prática da taxonomia inferir sobre o processo de identificação botânica.

A inclusão de vetores semânticos na prática de mensurar o grau de relevância das características botânicas de cada espécie permite uma maior exatidão nos resultados, além de quantificar por porcentagem o grau de semelhança entre as informações do usuário e as

espécies catalogadas, fornecendo, desta forma, indicativos mais detalhados para a tomada de decisão.

O sistema especialista para classificação de imagens de madeira apresenta um método de reconhecimento de padrão proposto com segmentação em matriz de coocorrência. Extração de características com o descritor de textura de Haralick e posterior classificação com Redes Neurais Artificiais chegaram a resultados satisfatórios para o reconhecimento das espécies amazônicas apresentadas na base de imagem.

Observa-se, neste trabalho, a influência da aplicação das lixas nos resultados das imagens e conseqüentemente nos resultados do sistema de reconhecimento de padrões de madeiras da Amazônia, evidenciando que, quanto mais polida for a madeira, maior a perícia dos resultados.

Conclui-se, desta forma, que o tratamento da superfície da amostra da madeira permite não somente a retirada de defeitos externos da madeira como aumentam a nitidez, brilho e luminosidade. Tais benefícios permitem uma maior exatidão na extração de padrões (características) que identifiquem uma espécie, o que repercute na assertividade da taxa de reconhecimento.

A integração entre a ontologia Onto-AmazonTimber e o sistema especialista permite uma formalização do conhecimento contínua, propiciando novos cenários de identificação botânica e formas de aplicações no âmbito da botânica.

Além disto, a Onto-TimberAmazon pode contribuir fortemente no âmbito ambiental com potenciais sugestões para: (i) o aprimoramento do desenvolvimento sustentável no manejo florestal no setor madeireiro; (ii) a contribuição na conservação ambiental e o desenvolvimento social na região amazônica; e (iii) uma melhor fiscalização no setor madeireiro.

Isto posto, o presente trabalho conclui seu objetivo de dispor um referencial semântico que aporte ao processo de identificação botânica de espécies florestais amazônicas, contribuindo para formalização e armazenamento do conhecimento de taxonomistas e mateiros na tarefa de identificação botânica, além de propiciar ferramentas que auxiliem a tomada de decisão no processo de identificação botânica de espécies florestais comercializadas na Amazônia.

6.2 Contribuições da Tese

Dentre as contribuições apresentadas no trabalho de tese destaca-se o desenvolvimento de um referencial semântico que aporte à produção de conhecimento no domínio da botânica mais especificamente no processo de identificação botânica. Durante o desdobramento do referencial semântico foi desenvolvido: (i) um modelo conceitual que aporte a um referencial semântico no âmbito da botânica, suportando processos de reconhecimento de padrões; (ii) uma ontologia Onto-AmazonTimber com relações semânticas e restrições axiomáticas que inferem sobre características e imagens botânicas dispostas hierarquicamente representadas por entidades e propriedades.

Outras contribuições foram produzidas em decorrência da elaboração e validação do modelo conceitual proposto, nomeadamente:

- Integração semântica entre o referencial semântico e o reconhecimento de padrões;
- Projeto e implementação de um Sistema Especialista para identificar espécies botânicas baseado em reconhecimento de padrões de características e imagens das espécies;
- Atribuição de Vetor Semântico como máquina de inferência para melhoria da classificação de padrões;
- Projeto e implementação de um Sistema Especialista para classificar imagens de madeira de espécies amazônicas;
- Desenvolvimento de algoritmos de reconhecimento de padrões de imagens e conhecimento de espécies amazônicas

Durante a vida útil da tese foram publicados os artigos mostrados na Tabela 20.

Tabela 20. Publicações

Nº	PUBLICAÇÃO	LOCAL	STATUS
1	PONTE, M. J. M.; LIMA, C. P. Botanical identification of Amazonian species through patterns recognition of wood and essential oil: A framework based on ontology. In: 2014 9th Iberian Conference on Information Systems and Technologies (CISTI), 2014, Barcelona. 2014 9th Iberian Conference on Information Systems and Technologies (CISTI). p. 1.	Conferência	Publicado
2	PONTE, M. J. M.; FIGUEIRAS P. A.; JARDIM-GONÇALVES R.; LIMA, C. P. Ontological interaction using JENA and SPARQL applied to Onto-AmazonTimber ontology. 7th IFIP WG 5.5/SOCOLNET Advanced Doctoral Conference on Computing, Electrical and Industrial Systems, DoCEIS 2016 in Technological Innovation for Cyber-Physical Systems - 7th IFIP WG 5.5/SOCOLNET Advanced Doctoral Conference on Computing, Electrical and Industrial Systems, DoCEIS 2016, Costa de Caparica, Portugal, April 11-13, 2016, Proceedings, VOLUME: 470, PUBLISHED: 2016 Advances in Information and Communication Technology ISSN1868-4238	Conferência / Revista Científica	Publicado
3	Sarraipa, J., Marcelino-Jesus, E., Oliveira, P., Amaral, P., Ponte, M., Costa, R., Zdravković, M. <i>Aquaculture Knowledge Framework</i> . In: Konjović, Z., Zdravković, M., Trajanović, M. (Eds.) ICIST 2016 Proceedings Vol.1, pp.227-234, 2016	Conferência	Publicado
4	PONTE, M. J. M.; FIGUEIRAS P. A.; BARATA L. E. S.; MOUTINHO V. P.; JARDIM-GONÇALVES R.; LIMA, C. P. Tecnologia Semântica: mais sustentabilidade para o manejo florestal na Amazônia. Revista da Madeira – REMADE issn: 0034-7582; ano 26, numero 149, p.52-53, Dez-2016	Revista Científica	Publicado

5	AMARAL P.; OLIVEIRA P.; PONTE M. J. M.; MATADO D.; COSTA R.; SARRAIPA J. SEMANTIC ANNOTATION OF AQUACULTURE PRODUCTION DATA. Proceedings of the ASME Congress 2016 - ASME International Mechanical Engineering Congress and Exposition - IMECE 2016 November 11-17, 2016, Phoenix, Arizona, USA	Conferência	Publicado
6	PONTE, M. J. M.; JARDIM-GONÇALVES R.; LIMA, C. P. Sistema de suporte à decisão no processo de identificação de espécies florestais na Amazônia. Revista Em foco. Issn: 1806-5864.Ed. 107; 2017.	Revista Científica	Publicado
7	PONTE, M. J. M.; VINENTE, J. F. V.; MOUTINHO, V. H. P.; BARATA, L. E. S.; JARDIM-GONÇALVES R.; LIMA, C. P. Identificação de madeiras amazônicas por reconhecimento de padrões utilizando processamento de imagens	Revista Científica	Submetido, aguardando resultado
8	PONTE, M. J. M.; JARDIM-GONÇALVES R.; LIMA, C. P. Onto-AmazonTimber: a Semantic Referential Supporting the Botanical Identification Process of Amazonian Species	Revista Científica	Submetido, aguardando resultado
9	PONTE, M. J. M.; FIGUEIRAS P. A.; COSTA R.; JARDIM-GONÇALVES R.; LIMA, C. P. Expert System for botanical identification of Amazonian species: an approach based on ontological interaction and Vector Semantic	Revista Científica	Em desenvolvimento
10	PONTE, M. J. M.; VINENTE, J. F. V.; JARDIM-GONÇALVES R.; LIMA, C. P. Integration of texture descriptors for pattern recognition of Amazonian forest species	Revista Científica	Em desenvolvimento

Ainda associados a esta tese, os trabalhos mostrados na Tabela 21 foram produzidos.

Tabela 21. Trabalhos acadêmicos

Tipo	Título	Autor	Status
Trabalho de conclusão de curso (TCC) – Curso Ciência da Computação – UFOPA.	Integração de descritores de textura (Haralick e coloração) para reconhecimento de padrão de espécies florestais da Amazônia	Julie Flávia Vieira Vinente	Em desenvolvimento
Iniciação Científica	Madeiras agrupadas comercialmente sob o mesmo nome vernacular no Oeste do Pará	Sávio Dill	Em desenvolvimento

6.3 Desafios Encontrados

A princípio, é importante considerar que mesmo grandes centros de pesquisa no Brasil com alta produtividade científica apresentam dificuldade em realizar pesquisa científica com qualidade. Isto ocorre em decorrência da escassez de recursos, incentivos, e por vezes, falta de mão-de-obra acadêmica. Tal situação é agravada na região Norte, posto que a distribuição espacial de instituições de pesquisa propiciam o isolamento com outros centros de pesquisa desfavorecendo a troca de experiência, parcerias e avanços tecnológicos.

No decorrer do desenvolvimento desta tese, foram identificados alguns pontos críticos no âmbito tecnológico e metodológico que podem auxiliar futuros pesquisadores interessados em imergir na problemática da identificação botânica e tecnologias semânticas. De se referir os seguintes:

- Aquisição de conhecimento: situa-se como ponto crítico no desenvolvimento do referencial semântico, visto que a qualidade da ontologia depende do sincronismo entre profundidade e abrangência do conhecimento angariado. A aquisição do conhecimento no âmbito da botânica apresenta agravante, a escassez de botânicos especialista em identificação botânica restringe a interação com estes profissionais. No entanto existem bases de dados, base de imagens, catálogos botânicos e parataxonomistas disponíveis para captação de conhecimento.
- Formalização do conhecimento: existem diversas ferramentas disponíveis para a formalização do conhecimento, contudo é necessário um profundo conhecimento da plataforma de trabalho e de uma metodologia de suporte. Além disso, é necessário uma contínua interação com especialista no do âmbito da problemática durante o decorrer da formalização do conhecimento, isto pode ocasionar morosidade e alto custo no desenvolvimento da ontologia.
- Integração entre ontologia e cenários de aplicação: no trabalho de tese foi utilizado a ferramenta protégé e a linguagem de programação orientada a objeto (java), foi utilizado a API Jena que trata com arquivos “.rdf” gerado pela ferramenta protégé. A interação ocorreu de forma satisfatória desde que as configurações pertinentes a API estejam bem definidas, no entanto quando se trata de dispositivos móveis existem APIs de acesso a ontologia específicas que apresentam limitações de desempenho, acesso e armazenamento.

6.4 Trabalhos Futuros

Como trabalhos futuros, espera-se a continuidade do desenvolvimento dos cenários de aplicação do referencial semântico, em destaque: SE para reconhecimento de espécies florestais por análise de imagens de carvão, SE para reconhecimento de espécies florestais por componentes químicos, aprimoramento do SE para reconhecimento de espécies florestais – inventário florestal e SE para reconhecimento de espécies florestais por imagem de madeira. Outro ponto trata-se do aprimoramento da ontologia Onto-AmazonTimber no que tange a atualização da base de conhecimento.

- SE para reconhecimento de espécies florestais por análise de imagens de carvão

Este cenário motiva-se pelo aumento da produção siderúrgica na Amazônia, o que vem exigindo um maior consumo de carvão como insumo energético, ocasionando o aumento das carvoarias, provocando, desta forma, um déficit ambiental. A obtenção de carvão proveniente de áreas das reservas florestais ou de determinadas espécies situa-se como crime ambiental, posto que não existe um retorno dos recursos florestais na produção do carvão.

Neste contexto, este cenário propõe um sistema computacional que objetiva identificar espécies botânicas por reconhecimento de padrões da imagem de carvão, obedecendo aos mesmos critérios, artifícios tecnológicos e epistemológicos do reconhecimento de madeira, visto que a imagem microscópica do carvão apresenta características provenientes da madeira original, compondo uma relação de madeira e carvão. Desta forma, as características anatômicas da madeira contidas no carvão situam-se como um identificador único, possibilitando sua padronização por espécie.

- SE para reconhecimento de espécies florestais por componentes químicos

Este SE objetivará reconhecer padrões de incidência e de concentração de substâncias químicas, contidas na madeira. Após análise química em laboratório os resultados serão analisados no sistema especialista que fornece regras para diagnóstico e classificação da espécie correspondente

- Aprimoramento do SE para reconhecimento de espécies florestais (inventário florestal)
 - Análise de localização automática via GPS, no intuito de verificar a distribuição e o domínio fitogeográfico da espécie identificada.
 - Inclusão de novas espécies florestais da Amazônia.
 - Inclusão de novas características que qualifiquem as espécies cadastradas.

- Inclusão de outros mecanismos que facilitem o diagnóstico da identificação das espécies.
- Aprimoramento do SE para reconhecimento de espécies florestais por imagem de madeira:
 - Validação do método de reconhecimento de padrão por imagem, a citar o método *leave one out cross validation*.
 - Inclusão de outros métodos para processamento de imagens de madeira: segmentação e classificação.
- Interface para gestão da ontologia

Como mencionado anteriormente, os domínios de conhecimento raramente são estáticos e, assim sendo, a ontologia deve acompanhar a evolução do domínio de conhecimento. Caso mudanças do domínio do conhecimento não sejam mapeadas e incorporadas na base de conhecimento, esta se tornará estagnada, ultrapassada, ineficaz e possivelmente incorreta. Desta forma, como trabalho futuro propõe-se uma interface para gestão (alteração, exclusão, inclusão) de conhecimento existente, perpassando por uma validação para compor a base de conhecimento.

- Melhorias com enriquecimento semântico por pesos em características florestais

Vislumbra-se outra perspectiva, com a observação de que cada espécie botânica tem particularidades quanto as características que podem diferenciá-las de outras espécies. Desta forma, sugere-se como trabalho futuro, melhorias na prática de mensurar por espécie a importância de cada característica botânica, que possibilite inserir uma interface para inclusão de pesos das características ou que possa ocorrer de forma automática. Para tal, pretende-se incluir vetores semânticos com pesos que qualifiquem a importância de cada característica para a espécie botânica analisada.

- Plataforma Móvel

A portabilidade e a conectividade dos dispositivos móveis integradas com os recursos computacionais de representação do conhecimento, permitindo, desta forma, um ambiente propício para o desenvolvimento de Sistemas Baseados em Conhecimento. Isto posto, os avanços tecnológicos provenientes das tecnologias móveis convergem para novos desafios no sentido de buscar o acoplamento do mundo físico ao mundo semântico, a fim de prover uma abundância de serviços e aplicações, possibilitando que software, hardware e usuário interajam de forma transparente.



7. REFERÊNCIAS BIBLIOGRÁFICAS

- AKSOY, S.; HARALICK, R. M. Using Texture in Image Similarity and Retrieval. Intl. Workshop on Texture Analysis in Machine Vision, pp 111-117, Finland, June.
- ALBINO, V.; GARAVELLI, A. C.; SHIUMA, G. A metric for measuring knowledge codification in organization learning. *Technovation*, n. 21, p. 413-422, 2001.
- ALMEIDA, M; BAX, M. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Revista Ciência da Informação*, 32(3), 2003
- AMIN, S. H. M.; ZAWAWI; A. A.; TIMAN, H. To share or not to share knowledge: Observing the factors. Conference Proceeding, IEEE Colloquium on Humanities, Science and Engineering, Malaysia. doi: 10.1109/CHUSER.2011.6163859, 2011.
- ANGELONI, M. T. (Coord.). Organizações do conhecimento: infraestrutura, pessoas e tecnologia. São Paulo: Saraiva, 2002.
- APACHE. Apache.org JENA. 2015. <http://jena.apache.org/>, Nov. 2015
- BAILÓN, B.A.; GIBAJA E., PÉREZ R. AND QUESADA C. G.R.E.E.N. - An Expert System to identify Gymnosperms. In *Proceedings of the Sixth International Conference on Enterprise Information Systems - Volume 2: ICEIS*, ISBN 972-8865-00-7, pages 216-221. In *Proceedings of the Sixth International Conference on Enterprise Information Systems - Volume 2: ICEIS*, 216-221, 2004, Porto, Portugal
- BARNESLEY M. F.; DEMKO S. "Iterated function systems and the global construction of fractals", *Proc. of Royal Soc. London*, vol. A399, pp. 243-275, 1985.
- BEAL, A. Gestão estratégica da informação: como transformar a informação e a tecnologia da informação em fatores de crescimento e de alto desempenho nas organizações. São Paulo: Atlas, 2004.
- BEER, S. Brain of the firm. 2nd ed. New York: John Wiley & Sons. 1981.

- BENNET, A.; BENNET, D. Organizational survival in the new world: the intelligent complex adaptive system. A new theory of the firm. Burlington, MA: Elsevier Science. 2004.
- BENNET, A.; BENNET, D. The fallacy of knowledge reuse: building sustainable knowledge. Journal of knowledge management, v. 12, n. 5, p. 21-33, 2008.
- BERNARAS, A.; LARESGOITI, I.; CORERA, J. Building and reusing ontologies for electrical network applications. In: THE EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE, ECAI, 1996. *Proceedings...* 1996. p. 298-302.
- BEZDEK, J. C., PAL, S. K. (Editores) Fuzzy Models Pattern Recognition: Methods That Search for Structures in Data, IEEE, 1992.
- BEZDEK, J. C., PAL, S. K. (Editores) Fuzzy Models Pattern Recognition: Methods That Search for Structures in Data, IEEE, 1992.
- BEYER, K.; GOLDSTEIN, J. R. Ramakrishnan, and U. Shaft. When is nearest neighbors meaningful? In Proc. International Conference on Database Theory, 1999.
- BHATT, G. D. Knowledge management in organizations: examining the interaction between technologies, techniques and people. Journal of Knowledge Management, 5 (1), 68-75, 2001.
- BIHUI W.; HENGNIAN Q. "Wood Recognition Based on Grey-Level Co-Occurrence Matrix." Paper presented at International Conference on Computer Application and System Modeling, Taiyuan, China, October 269–272, 2010.
- BISHOP, C.. Neural Networks for Pattern Recognition. Birmingham: Oxford University Press. 1995.
- BITTENCOURT, I. I., ISOTANI, S., COSTA, E. & MIZOGUCHI, R. Research Directions on Semantic Web and Education. Journal Scientia, 2008, P. 19(1), 59-66
- BISHOP, C.. Neural Networks for Pattern Recognition. Birmingham: Oxford University Press. 1995.
- BOOCH, Grady. "UML in Action", Communications of the ACM, vol 42, n 10, Oct 1999

- BRADSHAW, C. J. A.; SODHI, N. S.; BROOK, B. W. Tropical turmoil: a biodiversity tragedy in progress. *Frontiers in Ecology and the Environment*, Washington, DC, v. 7, n. 2, p. 79–87, 2009.
- BRAGA, A. P.; CARVALHO, A. C.; LUDEMIR, T. B. *Redes Neurais Artificiais*. In: REZENDE, Solange O. (Org.). *Sistemas Inteligentes: fundamentos e aplicações*. Barueri, SP: Manole, p.141-168, 2003.
- BRANTS T.; CHEN F.; TSOCHANTARIDIS I. Topic-based document segmentation with probabilistic latent semantic analysis. In *Conference on Information and Knowledge Management (CIKM)*, pages 211–218, 2002.
- BRANCH A. R., SONG B. H. ActKey: A Web-Based Interactive Identification Key Program. *International Association form Plant Taxonomy (IAPT)* Vol. 54, No 4, Nov. 2005. pp 1041 – 1046
- BRANCO, R.R. Ontologias para levantamento e avaliação de aspectos e impactos ambientais oriundos de empreendimentos de Engenharia: proposição de um modelo conceitual / (Dissertação) Mestrado em Engenharia Civil – Universidade Federal Fluminense, RJ, 2013.
- BREITMAN, Karin. *Web Semântica: a Internet do futuro*. Rio de Janeiro: LTC, 2005
- BORST W. Construction of Engineering Ontologies. PhD thesis, Institute for Telematica and Information Technology, University of Twente, Enschede, The Netherlands, 1998
- BROWN M.; GRUNDY W.; LIN D.; CRISTIANINI N.; SUFNET C.; FUREY T.; ARES J.; and HAUSSLER D. Knowledge-based analysis of microarray gene expression data by usinf support vector machines. *Proc. Natl. Acad. Sci. USA*, 97, 262 – 267. 2000.
- BUNGE, M. Emergence and convergence: qualitative novelty and the unity of knowledge. Toronto: University of Toronto, 2003. 330 p.
- CAMARGOS, J.A A. et al. Catálogo de árvores do Brasil. Brasília: IBAMA (Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis); Laboratório de Produtos Florestais. p887,1996.

- CAMPOS, J. A. G.; Análise conceitual sobre as relações semânticas em ciência da informação: contribuições para o desenvolvimento de ontologias. Dissertação. Programa de Pós-Graduação em Ciência da Informação - Universidade Federal de Minas Gerais. Belo Horizonte, 2009.
- CAMPOS DOS SANTOS, J.L.; BY, R. A.; MAGALHÃES, C. “A Case Study of INPA's Bio-DB and an Approach to Provide an Open Analytical Database Environment”. *International Archives of Photogrammetry and Remote Sensing*, 33 (B4): 155-163, 2000.
- CAMPOS DOS SANTOS, J.L.; MAGALHÃES NETTO, J F., CASTRO, A. N.; ALBUQUERQUE A. C. F.; FERNEDA, E.; ALONSO, L.; ROCHA, R. L. C.; MENDONÇA, D.T. “Ontologias para Interoperabilidade de Modelos e Sistemas de Informação de Biodiversidade“. In: *Proceedings of the Iberoamerican Meeting of Ontological Research collocated CITA 2011*. Aachen: CEURWS.org. v.728. Gramado. 2011.
- CARVALHO, André; LUDEMIR, Antônio. *Fundamentos de Redes Neurais Artificiais*: 11ª Escola de Computação. Imprinta Gráfica e Editora Ltda, 1998.
- CASTLEMAN, K. R. *Digital Image Processing*. Upper Saddle River: Prentice Hall, Inc. 1996
- CASANOVA, D. Identificação de espécies vegetais por meio da análise de textura foliar. Dissertação de Mestrado; Instituto de Ciências Matemáticas e de Computação (ICMC); Universidade de São Paulo – USP, São Carlos, 2008
- CAVALCANTI, M. C. B.; GOMES, E. B. P.; NETO, A. P. *Gestão de empresas na sociedade do conhecimento: um roteiro para ação*. Rio de Janeiro: Editora Campus Eslevier, 2001.
- CHAIBEN, H. *Inteligência Artificial na Educação*. <
<http://www.cce.ufpr.br/~hamilton/iaed/iaed.htm>> Acesso em 05 ago. 2016
- CHENG, S.J., MALIK, O.P., HOPE, G.S. “An Expert System for Voltage and Reactive Power Control of a Power Systems” – *IEEE Trans. On Power Systems*, vol. 3, no.4, nov/1988, pp. 1449 – 1455.
- CHOO, C. *The knowing organization*. New York: Oxford University Press. 1998.

- CHOO, C. W. A organização do conhecimento: como as organizações usam a informação para criar significados, construir conhecimento e tomar decisões. Tradução de Eliana Rocha. São Paulo: Senac São Paulo, 2006.
- CONAMA, BRASIL, República Federativa. Resolução CONAMA nº 001, de 23 de janeiro de 1986.
- CONNERS, R. W. A Theoretical Comparison of Texture Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1980. 2:204-222.
- CONNERS, R. W.; TRIVEDI, M. M. Segmentation of a High-Resolution Urban Scene using Texture Operators. *Computer Vision, Graphics and Image Processing*. 1984. 25:273-310
- CORDEIRO, F. M. Reconhecimento e Classificação de Padrões de Imagens de Núcleos Linfócitos do Sangue Periférico Humano com a Utilização de Redes Neurais Artificiais. Universidade Federal de Santa Catarina, 2002.
- CORDINGLEY, E. Knowledge elicitation techniques for knowledge-based systems. In: DIAPER, D. (Ed.) *Knowledge Elicitation: Principles, Techniques and Applications*. Chichester: Ellis Horwood Limited, pl 89-178, 1989.
- COSTA, R. D. D. Semantic Enrichment of Knowledge Sources Supported by Domain Ontologies. Tese de Doutorado em Engenharia Electrotécnica e de Computadores – Faculdade Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal - Lisboa, 2014.
- CUNHA, F.S. Um Sistema Especialista para Previdência Privada. Florianópolis, 1995. Dissertação (Mestrado em Engenharia de Produção) – Departamento de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, [s.p.] Disponível em: <http://www.eps.ufsc.br/disserta/cunha/indice/index.html>. Acessado em dez 2015
- DALKIR, K. *Knowledge Management in Theory and Practice*, Burlington: Elsevier Butterworth-Heinemann. 2005.
- DAUGMAN J. G. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, 2(7):1160–1169, July 1985.

- DAUGMAN, J. D. Complete discrete 2-d Gabor transforms by neural networks for image analysis and compression. Em: IEEE Trans. Acoustics, Speech and Signal Processing. V. 36, p. 1169 - 1179, 1988
- DAVENPORT, T.H.; PRUSAK, L. Working knowledge: How organizations manage what they know. Harvard Business School Press, Boston. 1998.
- DAYHOFF, J. E. Neural Networks Architectures, New York: Van Nostrand Reinhold, 1992.
- DE PAULA, P. L. F.; OLIVEIRA, L.S.; NISGOSKI, S.; BRITTO, A.S.J. Forest species recognition using macroscopic images. Springer-Verlag Berlin Heidelberg - Machine Vision and Applications 25:1019–1031, 2014.
- DE PAULA, P.L.F. Reconhecimento de espécies florestais através de imagem macroscópicas. Tese de doutorado. Programa de Pós- Graduação em Informática do Setor de Ciências Exatas da Universidade Federal do Paraná, 2012.
- DE PAULA, P.L.F.; TUSSET, A.M. Análise de cor para reconhecimento de espécies florestais. I SIPEX – Seminário Integrado de Pesquisa e Extensão Universitária. Editora: ÁGORA: revista de divulgação científica, v. 16, n 2(A), 2009.
- DEERWESTER S.; DUMAIS S.; FURNAS G.; LANDAUER T.; HARSHMAN R. Indexing by latent semantic analysis. Journal of the American Society for Information Science, 41(6):391–407, 1990.
- DEZA, M. M.; DEZA, E. *Encyclopedia of Distances*. Heidelberg: Springer-Verlag Berlin Heidelberg. 2009.
- DHAR, V.; STEIN, R. Seven methods for transforming cooperate data into business intelligence, Prentice Hall International, New Jersey, 1997, 270 p. il. ISBN 0-13-282006-4.
- DOD. Documentation DoD Architecture Framework, Versão 1.0 DoD 5000.59-P, "Modeling and Simulation Master Plan,". IEEE STD 610.12. / IAP (Integrated Architectures Panel) / C4ISR ITF (Integration Task Force). 1995

- DRUMOND, L.; GIRARDI, R. Extracting ontology concept hierarchies from text using markov logic. In: SYMPOSIUM ON APPLIED COMPUTING, 25., 2010, Switzerland. Proceeding Switzerland, 2010.
- DUDA, O., HART, P. E. Pattern classification and scene analysis. John Wiley & Sons, Inc., 1973.
- DUBIN, D. The Most Influential Paper Gerard Salton Never Wrote. 2004. *Library Trends*. 52(4), pp. 748-764
- FERNÁNDEZ-LÓPEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, N. Methontology: From ontological art towards ontological engineering. Paper presented at the Spring Symposium on Ontological Engineering of AAAI (pp. 33-40). Stanford University, 1997.
- FIALHO et al. Gestão do Conhecimento e aprendizagem: as estratégias competitivas da sociedade pós-industrial. Florianópolis: Visual Books, 2006.
- FIGUEIREDO, S. P. Gestão do conhecimento: estratégias competitivas para a criação e mobilização do conhecimento na empresa: descubra como alavancar e multiplicar o capital intelectual e o conhecimento da organização. Rio de Janeiro: Qualitymark, 2005.
- FIRESTONE, J.M.; McELROY, M.W. Key issues in The New Knowledge Management. Amsterdam: KMCI/Butterworth-Heinemann, 2001.
- FOGEL, I.; SAGI, D. (1989). "Gabor filters as texture discriminator". *Biological Cybernetics*. 61 (2). doi:10.1007/BF00204594. ISSN 0340-1200.
- FORTES, L. F.; LABIDI S.; CARVALHO, R. F. Projeto e construção de bio-ontologia para suporte a sistema multiagente de monitoramento ambiental do complexo portuário da ilha de São Luís do Maranhão. *Revista Científica – Caderno de Pesquisa*. 2008.
- FU, K. S.; MUI, J. K. A survey on image segmentation. *Pattern Recognition* , v. 13, p. 3–16, Ezmsford, 1981.
- GABOR, D. "Theory of communication", *Journal of IEEE*. 1946. v. 93, p. 429459.

- GAGVANI, N. “Introduction to video analytics.” [S.I]. 2008. Disponível em: “<http://www.eetimes.com/design/industrial-control/4013494/Introduction-to-video-analytics>”. Acesso em 14 jan. 2011.
- GALLIANO. O método científico: teoria e prática. São Paulo: Harbra, 1979.
- GARCIA, A.C. B.; VAREJÃO, F.M.; FERRAZ, I. N. Aquisição de conhecimento. In: REZENDE, Solange O. (Coord.). Sistemas inteligentes: fundamentos e aplicações. São Paulo: Manole, 2005.
- GENARO, S. Sistemas especialistas: o conhecimento artificial. São Paulo: Editora S.A., 1986
- GIL, A. C. Métodos e Técnicas de Pesquisa Social. 6 ed. São Paulo: Atlas, 2008
- GOLDSTEIN, J. E RAMAKRISHNAN, R. Contrast plots and p-sphere trees: space vs. time in nearest neighbor searches. In Proc. 26th International Conference on Very Large Databases, pages 429-440, 2000.
- GÓMEZ-PÉREZ, A.; FERNÁNDEZ-LÓPEZ, M.; CORCHO, Oscar. Ontology Engineering – with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. London: Springer-Verlag, 2004.
- GONZALEZ R, WOODS R. Digital image fundamentals . Digital Image Processing. 2 ed. New Jersey: Prentice; 2002 :34-75.
- GROSAN, C.; ABRANHAM, A. Intelligent Systems: A modern approach. Book Intelligent Systems. Vol 17. 2011
- GRÜNINGER, M.; FOX, M.S. Methodology for the design and evaluation of ontologies. In D. Skuce (Ed.), IJCAI95 Workshop on Basic Ontological Issues in Knowledge Sharing (pp. 6.1–6.10).1995.
- GRUBER, T.R. A Translation Approach to Portable Ontologies. In: *Knowledge Acquisition*, v. 5, n. 2, p. 199-220, 1993
- GUARINO, N. Formal ontology and information systems. Em: Proceedings of the 1st International conference on formal ontologies in information systems, Trento, Itália. Amsterdã: IOS Press, 1998.

- GUARINO, N., OBERLE, D., STAAB, S. (2009). What is an Ontology? In: Staab, S., Studer, R. (eds.) *Handbook on Ontologies*, 2nd edn., pp. 1–17. Springer, Heidelberg. 2009.
- GUPTA, A. K.; McDANIEL, J. Creating competitive advantage by effectively managing knowledge: a framework for knowledge management. *Journal of Knowledge Management Practice*, 2002.
- GU, F.; CAO, C.; SUI, Y.; TIAN, W. Domain-Specific Ontology of Botany. *J Computer Science & Technology*. Vol. 19, No.2, p. 238-248, 2004.
- HAIPENG Y.; YIXING, L.; ZHENBO, L. “Wood Species Retrieval on Base of Image Textural Features” *Scientia Silvae Sinicae* 43(2007): 77–81.
- HANGJUN W.; HENGNIAN Q.; “Segmentation Method of Wood Microscopic Image Based on Local Level Set.” *Journal of Image and Graphics* 17: 236–240, 2012-1.
- HANGJUN W.; HENGNIAN Q.; XIAOFENG W. “A New Wood Recognition Method Based on Gabor Entropy.” Paper presented at *Advanced Intelligent Computing Theories and Applications: With Aspects of Artificial Intelligence*, Changsha, China, August 435–440, 2012-2.
- HANGJUN W.; BIHUI W. “A Novel Method of Softwood Recognition.” *Scientia Silvae Sinicae* 47: 141–145, 2011.
- HANGJUN W.; HENGNIAN Q.; WENZHU, L.; GUANGQUN Z.; PAOPING W. “A GA-Based Pore Feature Automatic Extraction Algorithm.” Paper presented at *In Proceedings of the First ACM/SIGEVO Summit on Genetic and Evolutionary Computation*, Shanghai, China, June 985–988, 2009.
- HARALICK, R.M. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786-804, 1979.
- HARALICK R., SHANMUGAN K., e DINSTEIN I., “Textural features for image classification,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.
- HAYKIN S., *Neural Networks and Learning Machines* (3rd Edition), Prentice Hall, 2009

- HAYES-ROTH, I et al. Building experts systems. Reading: Addison-Wesley, 1983, 444p
- HENGNIAN Q.; FENG NONG C.; HANGJUN W. “Analysis of Quantitative Pore Features Based on Mathematical Morphology.” *Forestry Studies in China* 10: 193–198, 2008.
- HEARST M.; SCHÜTZE H. Customizing a léxicon to better suit a computational task. In *ACL SIGLEX Workshop*, Columbus, Ohio, 1993.
- HEINZLE, R. Protótipo de uma ferramenta para criação de sistemas especialistas baseados em regras de produção. Florianópolis : UFSC, 1995. Dissertação de Mestrado, Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Engenharia de Produção e Sistemas.
- HOPLIN, H.P.; ERDMAN, S. J. Expert Systems: An Expanded Field of Vision for Business. In: Awad EM (ed) *Conference on Trends and Directions in Expert Systems*. ACM Press, New York, NY, 1990. pp 1-16
- HORROCKS, I.; PATEL - SCHNEIDER, P. F.; HARMELEN, F. V. From SHIQ and RDF to OWL: the making of a Web Ontology Language. *Web Semantics: Science, Services and Agents on the World Wide Web*, v. 1, n. 1, p. 7 - 26. doi: DOI: 10.1016/j.websem.2003.07.001, 2003.
- HORRIDGEN, KNUBLAUCH, H, RECTOR A. A practical guide to building OWL ontologies using the protégé-OWL plugin and CO-ODE tools edition 1.0 *The University of Manchester and Stanford University*. 2004. pp 126-141
- HUI W.; YAN L.; KAI Z. “Research on Color Space Applicable to Wood Species Recognition” *Forestry Machinery & Woodworking Equipment* 37: 20–22, 2009.
- HWA, L.K. “Planejamento da Expansão a Longo Prazo de Redes de Transmissão de Energia Elétrica Usando Técnicas de Sistemas Baseados em Conhecimento” – Tese de mestrado, COPPE-UFRJ, jul/1987
- ICMBIO <http://www.icmbio.gov.br/flonatapajos/mapas-e-limites.html>. Acesso: jan/2017
- ISO/IEC 10746-2: Information Technology - Open Distributed Processing - Reference Model: Foundations

- JUN T. B.; KIM D. A compact local binary pattern using maximization of mutual information for face analysis, *Pattern Recognition* 44 (3 (March)) (2011) 532–543
- JAIN, A. K.; DUIN, R. P.W.; MAO, J. Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4- 37, January, 2000.
- JURAFSKY D.; MARTIN J.H . *Speech and Language Processing*. Copyright 2015. Allrights reserved. Cap 19 – Vectors Semantics. Draft of August 24, 2015.
- JINJIANG L., DA YUAN, Q. X.; ZHANG C., "Fractal Image Compression by Ant Colony Algorithm", *Young Computer Scientists 2008. ICYCS 2008. The 9th International Conference for*, pp. 1890-1894, 2008.
- JONES,D;BENCH-
CAPON,T;VISSER,P.Methodologiesforontologydevelopment.inProc.ITKNOWS
Conference,XVIFIPWorldComputerCongress,Budapest,August., 1998
- KALABOKIDIS, K.; ATHANASIS N.; VAITIS, M. OntoFire: an ontology-based geo-portal for wildfires in *Nat. Hazards Earth Syst. Sci.*, 11, 3157–3170, 2011.
- KATAMAYA, N. E SATOH, S.The SR-tree: An index structure for high-dimensional nearest neighbor queries. In *Proc. 1997 ACM SIGMOD International Conference on Management of Data*. 1997.
- KAUPPINEN T.; ESPINDOLA G.M. OntologyBased Modeling of Land Change Trajectories in the Brazilian Amazon. In *Geoinformatik 2011—GeoChange*, Münster, Germany,, June 2011.
- KHALID M., LEE, E.L.Y., YUSOF, R., NADARAJ, M. Design Of An Intelligent Wood Species Recognition System. *IJSSST*, Vol. 9, No. 3, September 2008.
- KHOO, C. S. G.; NA, Jin-Cheon. Semantic relations in Information Science. *Annual Review of Information Science and Technology*, v. 40, p. 157-228, 2006
- KOHONEN, Teuvo. *Self-Organization and Associative Memory*. Springer-Verlag Series in Information Science. 1987.

- KROGH, A.; VEDELSBY, J.. *Neural Network Ensembles, Cross Validation and Active Learning* in NIPS - Advances in Neural Information Processing Systems, Vol. 7, pp. 231-238, The MIT Press, 1995.
- KURTZ, C. ; GANÇARSKI, P. ; PASSAT, N. ; PUISSANT, A.: A hierarchical semantic based distance for nominal histogram comparison. En: Data & Knowledge Engineering 87, p. 206–225. 2013.
- KVALHEIM, O. M. Interpretation to latent-variable for projection methods and their use and aims in the interpretation of multicomponent spectroscopic and chromatographic data. *Chemometrics and Intelligent Laboratory Systems*, 4: 11-15. 1998.
- LAKATOS, Eva Maria; MARCONI, Marina de Andrade. *Metodologia científica*. São Paulo: Atlas, 1983.
- LANDAUER T. e DUMAIS S. A solution to Plato's problem: The latent semantic analysis theory of acquisition. *Psychological Review*, 104(2):211–240, 1997.
- LAVALLE, M.M., RODRIGUEZ, G. “Análisis de Herramientas para Desarrollo de Sistemas Expertos” *Boletín IIE*, mayo/junio 1989, vol.13, no 3.
- LA SALLE, A. J.; MEDSKER, L. R. The wxpert system cycle: what have we learned from software engineering? *SIGBDP '90 Proceedings of the 1990 ACM SIGBDP conference on Trends and directions in expert systems*. Florida- USA 1990. pp 17-26
- LE COADIC, Y.F. *A ciência da informação*. Brasília, DF, Briquet de Lemos. 2004.
- LEE, C. J.; WANG, S. D. A Gabor filter-based approach to fingerprint recognition. In: *IEEE WORKSHOP ON SIGNAL PROCESSING SYSTEMS - SiPS*, 1999. Proceedings... . [S.l.]: IEEE, 1999. p. 371-378.
- LEGG, Catherine. Ontologies on the Semantic Web. *Annual Review of Information Science and Technology*, 2007, p. 407-451.
- LEFF, E. *Epistemologia ambiental*. 2. ed. São Paulo: Cortez, 2002.
- LEVENBERG, K. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Math*, Providence, v. 2, p. 164-168, 1944.

- LIAO, P. S., CHEN, T. S., e CHUNG, P. C. (2001). A fast algorithm for multilevel thresholding. *J. Inf. Sci. Eng.*, 17(5), 713-727. 2001.
- LIMA, L.C. “Sistemas Especialistas Aplicados ao Processamento de Alarmes em Centros de Controle” – Tese de mestrado, COPPE-UFRJ, jan/1988.
- LIMA, C., DIRABY, T., FIES, B., ZARLI, A., AND FERNELEY, E. The E-Cognos Project: Current Status and Future Directions of an Ontology-Enabled IT Solution Infrastructure Supporting Knowledge Management in Construction. *Construction Research Congress*, 2003, pp. 1-8.
- LIMA, C.; EL-DIRABY, T.; STEPHENS, J. Ontology-based optimisation of knowledge management in e-Construction. *ITcon* 10, 305–327. 2005.
- LISBOA F., J. Modelagem de Banco de Dados Geográficos. In: LADEIRA, M.; NASCIMENTO, M. E. M. III Escola Regional de Informática do Centro-Oeste. Brasília – DF. SBC – Sociedade Brasileira de Computação. 2000.
- LINJIN M.; HANGJUN W. “A New Method for Wood Recognition Based on Blocked HLAC.” Paper presented at Eighth International Conference on Natural Computation, Chongqing, China, May 29–31, 2012.
- LINGJUN S.; HANGJUN W.; HENGNIAN Q.; “Wood Recognition Based on Block Local Binary Pattern (LBP).” *Journal of Beijing Forestry University* 33: 107–112, 2011-1.
- LINGJUN S.; ZHIWEI J.; HANGJUN W. “A New Wood Recognition Method Based on Texture Analysis.” *Applied Mechanics and Materials* 58-60: 613–617, 2011-2.
- LNCC – Laboratório Nacional de Computação Científica – Ministério da Ciência e Tecnologia, acesso em: outubro de 2008. http://www.lncc.br/~labinfo/tutorialRN/frm4_backpropagation.htm
- LUDWIG, J. A.; REYNOLDS J. F. *Statistical Ecology. A primer on Methods and Computing* j. Wiley and Sons, Inc., New York. 1988.
- LUCKESI, C. C. e PASSOS, E.S. *Introdução à filosofia: aprendendo a pensar*. São Paulo: Cortez, 1996.

- LUGER, G. F. Artificial Intelligence: Structures and Strategies for Complex Problem Solving, 6th ed. Boston: Addison-Wesley Pearson Education, 2009.
- LUND K. e BURGESS C. Producing high-dimensional semantic spaces from lexical co-occurrence. Behavior research methods, instruments and computers, 28(22): 203–208, 1996.
- MA, Y. ; GU, X. ; WANG, Y.: Histogram similarity measure using variable bin size distance. En: Computer Vision and Image Understanding 114, p. 981–989. 2010.
- MAEDCHE. A Ontology Learning for the Semantic Web. Kluwer Academic Publishers, 2002.
- MALLIK A.; TARRÍO-SAAVEDRA J.; FRANCISCO-FERNÁNDEZ M.; NAYA S.; “Classification of Wood Micrographs by Image Segmentation.” Chemometrics and Intelligent Laboratory Systems 107: 351–362, 2011.
- MANNING, C.; PRABHAKAR R.; HINRICH S. Introduction to Information Retrieval. New York, NY: Cambridge University Press, 2009.
- MARTINS J.; OLIVEIRA L.; SABOURIN R. Combining textural descriptors for forest species recognition. In Proceedings of the 38th Annual Conference on IEEE Industrial Electronics Society. 1483—1488, 2012.
- MARTINS-DA-SILVA, R.C.V. *Coleta e Identificação de Espécimes Botânicos*. Embrapa Amazônia Oriental, Doc. 143. Belém. 40p. 2002
- MASTELA, L. S. Técnicas de Aquisição de Conhecimento para Sistemas Baseados em Conhecimento. Disponível em: <http://www.inf.ufrgs.br/gpesquisa/bdi/publicacoes/files/TI1LSM.pdf>. Acesso em 30/6/2004.
- MARCELJA, S. (1980). "Mathematical description of the responses of simple cortical cells". Journal of the Optical Society of America. 70 (11): 1297–1300. doi:10.1364/JOSA.70.001297.

- MARCHIORI, J.N.C., MUÑIZ, G.I.B. Estudo anatômico da madeira de *Myrceugenia glaucescens* (Camb.) Legr. et Kaus. Ciência e Natura, Santa Maria, 1988. v. 10, p. 105-113.
- MARENGONI, M.; STRINGHINI, S. Revista de Informática Teórica e Aplicada. V. 16, n.1, 2009
- MARQUARDT, D. W. An algorithm for least-squares estimation of nonlinear parameters. Journal of the Society for Industrial and Applied Mathematics, Philadelphia, v. 11, n. 2, p. 431–441, 1963.
- MARQUES, O. F.; VIEIRA, H. N. Processamento Digital de Imagem, Rio de Janeiro: Brasport. 1999.
- MENDES, R. D. Inteligência Artificial: Sistemas Especialistas no Gerenciamento da Informação.< <http://www.ibict.br/cionline/260197>> Acesso em 07 ago. 2016.
- MENESES, P. R., AND ALMEIDA, T. Introdução ao Processamento de Imagens de Sensoriamento Remoto. Brasília: Universidade de Brasília. 2012.
- MICROSOFT. Microsoft Product Documentation Visual Studio 6.0. What's in the Enterprise Edition.(2016)
- MIRANDA, R, C, da R. Gestão do Conhecimento Estratégico: uma proposta de modelo. Tese (Doutorado) – Faculdade de economia, administração, contabilidade ciência da documentação: Departamento de ciências da informação e documentação – UNB, Brasília, 2004.
- MINSKY, M. L.; PAPERT, S. A. *Perceptrons: An Introduction to Computational Geometry*, MIT Press, Cambridge, Massachusetts. 1969.
- MILLETTE, L., "Improving the Knowledge-Based Expert System Lifecycle" (2012).University of North Florida (UNF) Theses and Dissertations.Paper 407. <http://digitalcommons.unf.edu/etd/407>
- MIYAZAKI, F.A. Uma abordagem baseada em ontologias e conectores para a integração semântica de ferramentas de análise de expressão gênica. (Dissertação) Programa de Interunidades em Bioinformática – Universidade de São Paulo (USP). 2011

- MONTEIRO, F. S. Modelagem conceitual: a construção de uma ontologia sobre Avaliação do Ciclo de Vida (ACV) para fomentar a disseminação de seus conceitos. Bacharelado em Biblioteconomia (Monografia), *Universidade de Brasília (UNB)*, 2006.
- MOUTINHO, V.H.P.; SOUZA, M.A.R.; SILVA, S.S. da. Lisboa, P.L.B. Identificação e caracterização anatômica das espécies comercializadas como pau-mulato no Estado do Amapá. Setec/GEA. 50p, 2008.
- MORSE L. E.; BEAMEN J. H.; SHETLER S.G. A computer System for editing diagnostic keys for flora North America. International association for Plant Taxonomy IAPT. Vol. 17, No 5, pp479 – 483. 1968
- MUKHERJEE S.; TAMAYO P.; MESIROV J.; SLONIM D.; VERRI A. and POGGIO T. Support Vector Machine classification of microarray data. Technical Report CBCL Paper 182/AI Memo MIT. 1999.
- NASCIMENTO, J. P. R. DO; MADEIRA, H. M. F. & PEDRINI, H. Classificação de Imagens Utilizando Descritores Estatísticos de Textura in XI Simpósio Brasileiro de Sensoriamento Remoto, pp. 2099-2106. Belo Horizonte. 2003.
- NIEVOLA, J. C. Sistema Inteligente para Auxílio ao Ensino em Traumatologia Crânio-encefálica. Florianópolis, 1995. Tese (Doutorado em Engenharia Elétrica – Sistemas de Informação) – Grupo de Pesquisas em Engenharia Biométrica, Universidade Federal de Santa Catarina, 168p
- NILSSON, N.J. Principles of artificial intelligence. Palo Alto, CA: Tioga, 1980, 476 p
- NILSON, Neils S. Principles of Artificial Intelligence, Springer Verlag, Berlin, 1982
- NISSSEN, M.; KAMEL, M.; SENGUPTA, K. Integrated analysis and design of knowledge systems and processes. Information Resources Management Journal, Hershey, v. 13, n. 1, p. 24-43, Jan./Mar. 2000.
- NONAKA, I.; TAKEUCHI, H. The knowledge-creating company: how Japanese companies create the dynamics of innovation. New York: Oxford University Press. 1995.
- NONAKA, I.; TAKEUCHI, H. Criação de conhecimento na empresa: como as empresas japonesas geram a dinamica da inovacao. Rio de Janeiro: Elsevier, 1997.

- NOY, N.; MC GUINNESS, D. L. Ontology development 101: A guide to creating your first ontology. Relatório Técnico, Stanford University, 2001.
- NOY, N. F; FERGERSON, R. W.; MUSEN, M. A. The knowledge model of Protégé-2000: Combining interoperability and flexibility. In: International Conference on Knowledge /engineering and Knowledge Management (EKAW' 2000), 2TH, Juan-les Pins, France, 2000.
- ODDEN, W. e KVALHEIM, O.M. Application of multivariate modelling to detect hydrocarbon components for optimal discrimination between two source rock types. *Applied Geochemistry*, 15: 611-627. 2000.
- OJALA M. T.; PIETIKÄINEN, D. H. A comparative study of texture measures with classification based on featured distributions, *Pattern Recognition* 29 (1 (January)) (1996) 51–59.
- OJANSIVU V.; HEIKKILÄ J. Blur insensitive texture classification using local phase quantization, *Proceedings of the International Conference on Image and Signal Processing (ICISP 08)*, Springer Press, 2008, pp. 236–243.
- OTSU, N. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296), pp.23-27. 1975.
- OSORIO, Fernando Santos. *Um Estudo sobre Reconhecimento Visual de Caracteres através de Redes Neurais*. Dissertação de Mestrado, CPGCC, UFRGS, Porto Alegre - Brasil. Outubro 1991.
- PACHECO, M. A. C. Algoritmos Genéticos: Princípios e Aplicações. ICA: Laboratório de Inteligência Computacional Aplicada - www.ICA.ele.puc-rio.br Marco. Departamento de Engenharia Elétrica. Pontifícia Universidade Católica do Rio de Janeiro, 2002.
- PAO, Y. Adaptive Pattern recognition and neural networks. Addison-Wesley, 1989
- PEDRINI, H. & SCHWARTZ, W. R. *Análise de Imagens Digitais, Princípios, Algoritmos e Aplicações*. Editora Thomson Pioneira. 2008.

- PEREZ, A. A.; GONZAGA, A.; ALVES, J. M. Segmentation and analysis of leg ulcers color images, in: Proceedings of the International Workshop on Medical Imaging and Augmented Reality, 2001, pp. 262–266.
- POLANYI, M. The Tacit Dimension. Massachusetts: Doubleday & Co, 1983.
- PRAX, J. Y. Manager La connaissance dans l'entreprise. [S.I.]: INSEP, 1997.
- PRADO, E. V. “Sistema Especialista para dimensionamento e seleção de equipamentos para pre-processamento de café.” - Tese de doutorado – UFV, 2001
- PROCÓPIO L. C.; SECCO R. S. A importância da identificação botânica nos inventários florestais: o exemplo do “tauari” (*Couratari* spp. e *Cariniana* spp. - Lecythidaceae) em duas áreas manejadas no estado do Pará. Acta Amazonica. P31 – 44, Manaus. 2008.
- PROTÉGÉ PROJECT (2010), disponível em: <http://protege.stanford.edu>.
- PUTMAN J., HALL P. *Architecting with RM-ODP* PTR que referencia o ISO/IEC 10746-2: Information Technology - Open Distributed Processing - Reference Model: Foundations. 2001
- QUINELATO, B.T.; MORI, D.J.; ROSI, R.M. OntoSIGF – a geographic information system based on ontologies applied to forestry. *Ambiência - Revista do Setor de Ciências Agrárias e Ambientais* Vol. 4 - Edição Especial, Guarapuava-PR, p. 161-170, 2008.
- RAMOS, N. P.; LUCHIARI A. AGEITEC-Agência Embrapa de Informação Tecnológica; *Árvore do Conhecimento: Monitoramento Ambiental*; acesso: 2016 Fonte: http://www.agencia.cnptia.embrapa.br/gestor/cana-de-acucar/arvore/CONTAG01_73_71120
- RECTOR, M.; YUNES, E. Manual de semântica. Rio de Janeiro: Ao Livro Técnico, 1980
- REED, S. L.; LENAT, D. B. *Mapping ontologies into cyc*. 2002. Disponível em: http://www.cyc.com/doc/white_papers/mapping-ontologies-into-cyc_v31.pdf. Acesso em: nov. 2016.
- RENNOLLS, K. A Partial Ontology for Forest Inventory and Mensuration. 2nd International Workshop on Forest and Environmental Information and Decision Support Systems,

- FEIDSS'05 in the DEXA'05 Workshops, p. 679-683. IEEE Computer Society Press. ISBN 0-7695-2424-9. web-ref, 2005.
- RICH, E. e KNIGHT, K. Sistemas Especialistas. In:_____. Inteligência artificial. McGraw Hill Ltda. São Paulo. p.632–644. 1993.
- ROWLEY, J. The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of information Science*, v. 33, n. 2, p. 163-180, 2007.
- ROSSETTI A. G., MORALES A. B. T. O papel da tecnologia da informação na gestão do conhecimento. *C i. Inf.*, Brasília, v. 36, n. 1, p. 124-135, jan./abr. 2007
- ROSENFELD, A.; KAK, A. C. *Digital Picture Processing*. Academic Press, Volume 2, Second edition. 1982.
- RUBENSTEIN-MONTANO, B. et al. A systems thinking framework for knowledge management. *Decisions Support Systems Journal*, v. 31, n. 1, p. 5-16, 2001.
- RUSSELL, S. J. e NORVIG, P.. *Artificial Intelligence: A Modern Approach*. Prentice Hall, Saddle River, NJ - 2ª Edição – 2004.
- SALTON, G., WONG A., and YANG C. “A vector space model for automatic indexing.” *Communications of the ACM*, 1975: 613-620.
- SAHLGREN, M. The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces. Ph.D. Dissertation, Department of Linguistics, Stockholm University. 2006.
- SHADBOLT, N.; O'HARA, K.; CROW, L. The experimental evaluation of knowledge acquisition techniques and methods: history, problems and new directions. *International Journal of Human-Computer Studies*, v. 51, n. 4, p. 729-755, 1999.
- SCHERB, A.; KUEHN, V.; KAMMEYER, K.D.; Pilot Aided Channel Estimation for Short Code DSSSS. *Proc. IEEE. International Symposium on Spread Spectrum Technologies and Applications, ISSSTA*, (1):39-43, Setembro, 2002

- SCHULTE, R.P., LARSEN, S.L., SHEBLE, G.B., WRUBEL, J.N. - "Artificial Intelligence Solutions to Power System Operating Problems" - IEEE Trans. on Power Systems, vol. PWR-2, no. 4, nov/1987, pp. 920-926.
- SCHÜTZE H. Automatic word sense discrimination. *Computational Linguistics*, 24(1):97–124, 1998.
- SERVIN, G.. ABC of knowledge management. NHS National Library for Health, 2005.
- SHAOCHUN, Y.; KEQI, W.; TIANHONG, D.; XUEBING B. "Analysis of Wood Classification Using L*a*b* Color Space." *Forestry Machinery & Woodworking Equipment* 35(2007): 28–30.
- SINGH, A.; SOLTANI, E. Knowledge management practices in Indian information technology companies. *Total Quality Management*, v. 21, n. 2, p.145-157, Feb. 2010
- SILVA, J.N.M. *Manejo Florestal*. Embrapa Amazônia Oriental, 3ª. ed. Belém, p.49, 2001
- SILVA, L.A.E.; SAMPAIO, J.O.; DALCIN, D.; ZIMBRÃO, G.; SOUZA, J.M. Gestão do Conhecimento Taxonômico Aplicado na Conservação da Flora Brasileira. 2009 Acesso: http://www.lbd.dcc.ufmg.br/colecoes/e-science/2010/67361_1.pdf em 12/12/2014
- SILVA, S. S. Uma ontologia para interoperabilidade entre padrões de descrição de dados em biodiversidade ABCD e Darwin Core. (Dissertação) Mestrado em Ciência da Computação / Programa de Pós – Graduação em Ciência da Computação, Universidade Federal do ABC, 2014
- SIMON, H. Models of man: social and rational. New York: John Wiley & Sons. 1957.
- SISFLORA/PA – Relatórios: Extração e comercialização de toras de madeira nativa por Município. Acesso em: 14 de Maio de 2016. Disponível em: <<http://monitoramento.sema.pa.gov.br/sisflora/index.php/relatorios>>.
- SOUZA, D.R. de; SOUZA, A.L. de; LEITE, H.G.; YARED, J.A.G. Análise estrutural em floresta ombrófila densa de terra firme não explora, Amazônia Oriental. *Revista Árvore*. Vol. 30, nº1, p. 75-87. Minas Gerais: Viçosa, 2006.

- SOUTO, K.C. “Sistema Especialista com Lógica Nebulosa para o cálculo em tempo real de indicadores de desempenho e segurança na monitoração de usinas nucleares” – Tese de doutorado - COPPE-UFRJ, 2005.
- SOLANGE, O et.al. *Sistemas Inteligentes: fundamentos e Aplicações*. Manole, 2003.
- SOWA, J.F. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Pacific Grove: Brooks/Cole, 2000.
- SOUZA, A.; CORCHO, O.; VILCHES-BLÁZQUEZ, L.; SALLES, P. Brazilian Cerrado: a case study of linked geographical and statistical data applied to Ecology. *Semantic Web jornal - SWJ*, 2014.
- STAAB, S; SCHNURR, H; STUDER, R; SURE, Y. Knowledge processes and ontologies. *IEEE Intelligent Systems*, 16(1):26–34., 2001.
- STOLLENWERK, M. F. L. Gestão do conhecimento: conceito e modelos. In: BRASÍLIA, E. U. de (Ed.). *Inteligência Organizacional e competitiva*. Brasília: Editora UNB, 2001. P. 143 – 163.
- STUDER, R.; BENJAMINS, V. R.; FENSEL, D. Knowledge Engineering: Principles and Methods. *Data and Knowledge Engineering, IEEE Transactions on Data and Knowledge Engineering* v. 25, n. 1-2, p. 161-197, 1998.
- SVEIBY, K. E. *A Nova Riqueza das Organizações: Gerenciando e avaliando patrimônios de conhecimento*. Tradução: Luiz Euclides Trindade Frazão Flo. Rio de Janeiro: Campus, 1998.
- SUSTAETA, J.G., MUTH, M.M., GALLEGOS, A.A. “Sistemas Expertos para Supervision y Control de Procesos Industriales:Ejemplo de Aplicación em uma Planta Termoeléctrica”- *Boletín IIE*, Jul/ago 1989, vol. 12, no 4.
- SWARTOUT, B.; RAMESH, P.; KNIGHT, K.; RUSS, T. Toward distributed use of large-scale ontologies. In A. Farquhar, M. Gruninger, A. Gómez-Pérez, M. Uschold, & P. van der Vet (Eds.), *AAAI’97 Spring Symposium on Ontological Engineering* (pp. 138-148). Stanford University, 1997.

- TAKEUCHI, H.; NONAKA, I. Gestão do Conhecimento. Tradução de Ana Thorell. Porto Alegre: Bookman, 2008.
- TALUKDAR, S.N., CARDOZO, E., PERRY, T. - "The Operator's Assistant - An Intelligent, Expandable Program for Power System Trouble Analysis" - IEEE Trans. on Power Systems, vol. PWRS-1, no. 3, aug/1986.
- TAYLOR, T., LUBKEMAN, D. "Applications of Knowledge-Based Programming to Power Engineering Problems" – IEE Trans. on Power Systems, vol. 4, no. 1, feb/1989, pp345-352
- THEODORIDIS, S.; KOUTROUMBAS, K. "Pattern Recognition: Theory and Applications", 2nd edition, Academic Press, 2003.
- TERRA, J. C. C. Gestão do Conhecimento: o grande desafio empresarial. São Paulo: Negócio Editora, 2000.
- TOU, J. T., GONZALEZ, R. C. Pattern Recognition Principles. Addison-Wesley Publishing Company, Massachusetts, 1981.
- TOU, J. Y; LAU, P. Y.; TAY, Y. H. Computer Vision-based Wood Recognition System. Proceedings of International Workshop on Advanced Image Technology (IWAIT 2007), (pp. 197- 202). Bangkok, 2007.
- TUOMI, I. Data is more than knowledge: implications of the reversed knowledge hierarchy for knowledge management and organizational memory. Journal of management information systems, v.16, n. 3, p. 107-121, 1999.
- TURNEY, P.; PANTEL, P. From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence*. 2010. pp. 141-188.
- UMMINGER, B.; YOUNG, S. "Information Management for Biodiversity: a Proposed U.S. National Biodiversity Information Center". In: Reaka-Kudla, M.L.; Wilson, D.E. & Wilson, E.O. (eds.), Biodiversity II: Understanding and Protecting Our Biological Resources. Washington, D.C., Joseph Henry Press. p. 491-504, 1997.
- USCHOLD, M.; GRUNINGER, M. Ontologies: Principles, methods and applications. Knowledge Engineering Review, v. 11, p. 93–136, 1996.

- USCHOLD, M.; KING, M. Towards a methodology for building ontologies. In D. Skuce (Ed.), IJCAI'95 Workshop on Basic Ontological Issues in Knowledge Sharing (pp. 6.1-6.10), Montreal, Montreal, Canada. 1995.
- VARELA, F. Whence perceptual meaning? A cartography of current ideas. In F. Varela and J. P. Dupuy (Eds.). Understanding Origin: Scientific Ideas on the Origin of Life, Mind, and Society (A Stanford University Interational Symposium) Boston Studies Phil.Sci, Kluwer Assoc., Dordrecht, 1992.
- VELLASCO, M. M. Home Page "Redes Neurais". ICA: Núcleo de Pesquisa em Inteligência Computacional Aplicada – PUC – Rio. 2000, <http://www.ele.puc-rio.br/labs/ica/icahome.html>, consulta em 25/03/2008
- VON KROGH, G.; ROOS, J. Organizational epistemology. New York: St. Martin's Press. 1995.
- WATERMAN, D.A.: A Guide to Expert System, Reading, MA: Addison-Wesley Publishing Co., 1986.
- WIIG, K.M. Knowledge Management: an introduction and perspective. The Journal of Knowledge Management, vol. 1, n. 1, p. 6-14, September, 1997.
- WILSON, T.D. The nonsense of 'knowledge management'. Infomation Research, v.1, n.8 2002.
- YOUSSEF, Y. A. et. Al. O brasil e as práticas de gestão do conhecimento. 2006.
- WEICK, K. Making sense of the organization. Malden, MA: Basil Blackwell. 2001.
- WEISS, S. M.; KULIKOWSKI, C. A. Solução de problemas e consulta a especialista. In Guia Prático para Projetar Sistemas Especialistas. Livros Técnicos e Científicos Editora S.A. Rio de Janeiro. P1-5. 1988.
- WESZKA, A. A comparative study of texture measures for terrain classification. IEEE Trans SMC. 1976. 6:269-285
- WHELAN, P. F.; MOLLOY D. Machine Vision Algorithms in Java: Techniques and Implementation, Springer, London. 2000

- WIDDOWS, D. Geometry and Meaning. Stanford, California: CSLI publications.29:481–485.2004.
- WIDDOWS D.. Unsupervised methods for developing taxonomies by combining syntactic and statistical information. InProceedingsofHumanLanguagTechnology /NorthAmericanChapteroftheAssociationforComputational Linguistics, Edmonton, Canada, 2003
- WIDDOWS, D., e FERRARO, K. Semantic vectors: A scalable open source package and online technology management application. In Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008). 2008. , pp. 1183- 1190.
- WOOLLEY J. B.; STONE N. D. Application of Artificial Intelligence to Systematics; Systex – A Prototype Expert System for Species Identification. Oxford Journal – Systematic Biology. Vol. 3 Issu 3, pp 248 – 267. 1987
- W3C – World Wide WEB Consortium. eXtensible Markup Language (XML). 2014. Disponível em <<http://www.w3.org/XML/>>, 2014
- ZABOT, Joao Batista. Gestão do conhecimento: aprendizagem e tecnologia: construindo a inteligência coletiva. São Paulo: Atlas, 2002.
- XUEBING B.; KEQI W.; WANG HUI W. “Research on the classification of wood texture based on Gray LevelCo- occurrenceMatrix” Journal of Harbin Institute of Technology 37(2005): 1667–1670
- ZHIWEI J.; HANGJUN W.; TAO H.; JIANXIN Y. “A Morphological Feature Extraction Method of Wood Pores Based on an Improved Growing Region Algorithm.” Journal of Beijing Forestry University 33: 64–69, 2011.

8. APÊNDICES

8.1 Cooperação Internacional

A mobilidade internacional ocorreu mediante o Regime de Cotutela Internacional em decorrência da parceria firmada pela minuta do convênio/acordo acadêmico internacional entre UFOPA e a Universidade Nova de Lisboa (UNL). Tal regime tem como objetivo propiciar o intercâmbio acadêmico, bem como estabelecer e fortalecer relações com universidades estrangeiras, resultando na obtenção de titulação válida e reconhecida nas duas instituições.

Dentre as atividades desenvolvidas na UNL destaca-se a participação no projeto AQUASMART, desenvolvido no Grupo de Investigação sobre Interoperabilidade de Sistemas (GRIS) integrado no Instituto UNINOVA - centros de inovação tecnológica.

O projeto AQUASMART integra a “Estratégia de Crescimento Azul” da União Europeia firmado como estratégia para 2020 pela Comissão Europeia, objetivando o crescimento sustentável marinho e marítimo. Globalmente, quase metade do peixe consumido pelos seres humanos é produzida por fazendas de peixes. Prevê-se que a produção mundial aumente de 45 milhões de toneladas em 2014 para 85 milhões em 2030, tornando a indústria da aquicultura o setor de produção de alimentos que mais cresce no mundo. A União Europeia necessita de uma indústria de aquacultura inovadora para fazer face à procura crescente de produtos do mar e aumentar as suas reservas comerciais.

O AQUASMART aborda o problema do acesso global ao conhecimento, promovendo o intercâmbio e reutilização de dados entre empresas de aquicultura e suas partes interessadas. O projeto AQUASMART é impulsionado pela necessidade empresarial das empresas europeias de aquicultura, quando as empresas têm objetivos de negócio que não podem alcançar devido à falta de instrumentos que lhes permitam gerir e acessar conhecimentos globais e grandes dados, numa perspectiva multilingue, Multi-sectorial e transfronteiras.

Vale ressaltar que parte da teoria e desenvolvimento da tese foram aplicados no projeto AquaSmart. Os tópicos seguintes apresentam os produtos desenvolvidos durante o período de mobilidade internacional.

8.1.1 Base de conhecimento para aquicultura

A Estrutura Semântica é parte integrante da Gestão do Conhecimento, cujo objetivo é proporcionar recursos que possibilitem o acesso ao conhecimento. Dentre os diversos serviços inclui: taxonomia de conceitos, vetores semânticos e ontologias.

Durante a mobilidade desenvolveu-se a ontologia AquaSmart (Figura 60) que trata de uma ontologia no domínio da aquicultura que será utilizada para compor diversos serviços para melhorias da gestão, entre eles: levantamentos de empresas e identificação de doenças de peixes.

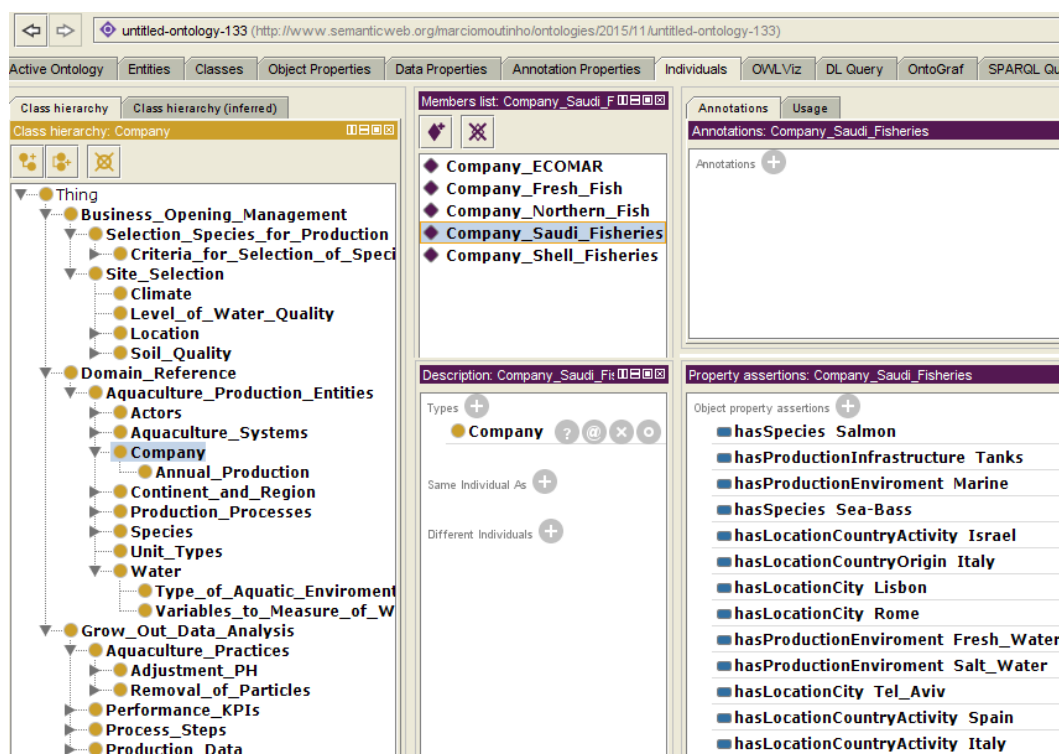


Figura 60 Ontologia AquaSmart

8.1.2 Sistema especialista para diagnóstico de doenças de peixes

Com a consolidação da aquicultura, emergem novas tecnologias de produção intensiva e a diversidade de espécies de peixes com potencial de cultivo, no entanto problemas sanitários ou de transmissão de doenças podem apresentar obstáculos nas diferentes fases de criação. Assim, o diagnóstico precoce de doenças e o manejo adequado constitui-se em fator primordial para o sucesso da atividade.

Neste sentido, o sistema de diagnóstico de doenças trata de um sistema computacional (Figura 61) que objetiva identificar doenças nas espécies de peixes, baseando-se em características físicas e comportamentais observadas na espécie e no ambiente em que está

inserida, isto inclui qualidade da água, presença de corpos estranhos ao ambiente de produção, como sedimentos e algas entre outras características.

Este serviço atua como interface do referencial semântico AquaSmart, no qual permite o diagnóstico precoce de doenças e sugestões de tratamento. Contudo o serviço não objetiva a substituição do veterinário, e sim propor práticas para o tratamento imediato e contenção de mortalidade, evolução e disseminação da doença.

É notório que o melhor controle das doenças é a prevenção com a adoção de boas práticas de manejo sanitário durante toda fase de produção. Mas, após a doença instalada, deve ser diagnosticada e tratada para prevenir grandes prejuízos.

AQUASMART DISEASE DIAGNOSIS	
<p>Inform fish species:</p> <p>Sea-Bass</p> <p>which syndromes?</p> <p> <input type="checkbox"/> None <input type="checkbox"/> Poor_growth <input type="checkbox"/> Immobility_with_the_head_down <input type="checkbox"/> Skin_cloudiness <input type="checkbox"/> Reduced_growth_rate <input type="checkbox"/> Disoriented_spinning_motion <input type="checkbox"/> Miliary_lesions_of_spleen_or_spleen_granulomatosis <input type="checkbox"/> Skin_area_haemorrhages <input type="checkbox"/> Lethargy <input type="checkbox"/> Hyperpigmentation <input type="checkbox"/> Distend_abdomen <input type="checkbox"/> Larvae_in_coelomatic_cavity <input type="checkbox"/> Ulcerations <input type="checkbox"/> Gill_lesions <input type="checkbox"/> Reduced_production <input type="checkbox"/> Low_mortality <input type="checkbox"/> Splenomegaly <input type="checkbox"/> Haemorrhages <input type="checkbox"/> Skin_ulcers <input type="checkbox"/> Necrosis <input type="checkbox"/> Fin_erosion <input type="checkbox"/> Gills_and_skin_tissue_necrosis <input type="checkbox"/> Necrotic_enteritis <input type="checkbox"/> Visceral_petechiation <input checked="" type="checkbox"/> Dark_skin <input type="checkbox"/> Anorexia <input type="checkbox"/> Focal_necrosis_of_the_gills <input type="checkbox"/> Skin_dusty_appearance <input type="checkbox"/> Focal_reddening_with_excess_mucus_production <input type="checkbox"/> Adults_and_larvae_on_fish <input type="checkbox"/> Hypertrophic_kidney_and_spleen_with_granulomas <input type="checkbox"/> Epithelial_hyperplasia <input type="checkbox"/> Gill_haemorrhages <input type="checkbox"/> Whitish_pseudotumor <input type="checkbox"/> Nervous_symptoms <input type="checkbox"/> Miliary_nodules_on_skin_or_gills <input type="checkbox"/> Skin_lesions <input type="checkbox"/> Depigmentation <input type="checkbox"/> White_spot_or_multifocal_white_patches <input type="checkbox"/> Emaciation <input type="checkbox"/> Head_down_position <input type="checkbox"/> Belly_up_syndrome </p> <p>SUBMIT</p>	<p>RESULTS</p> <p>Vibriosis_Sea-Bass</p> <p>Amyloodiniasis</p> <p>DISEASE TYPE: <i>Dinoflagellate</i> DISEASE AGENT: <i>Amyloodinium_occelatum</i> SUGGESTED MEASURES: <i>Freshwater treatment</i></p> <p>Photobacteriosis</p> <p>DISEASE TYPE: <i>Bacterium</i> DISEASE AGENT: <i>Photobacterium_damselae_subsp_pasteurella</i> SUGGESTED MEASURES: <i>Antibiotic treatment</i></p>

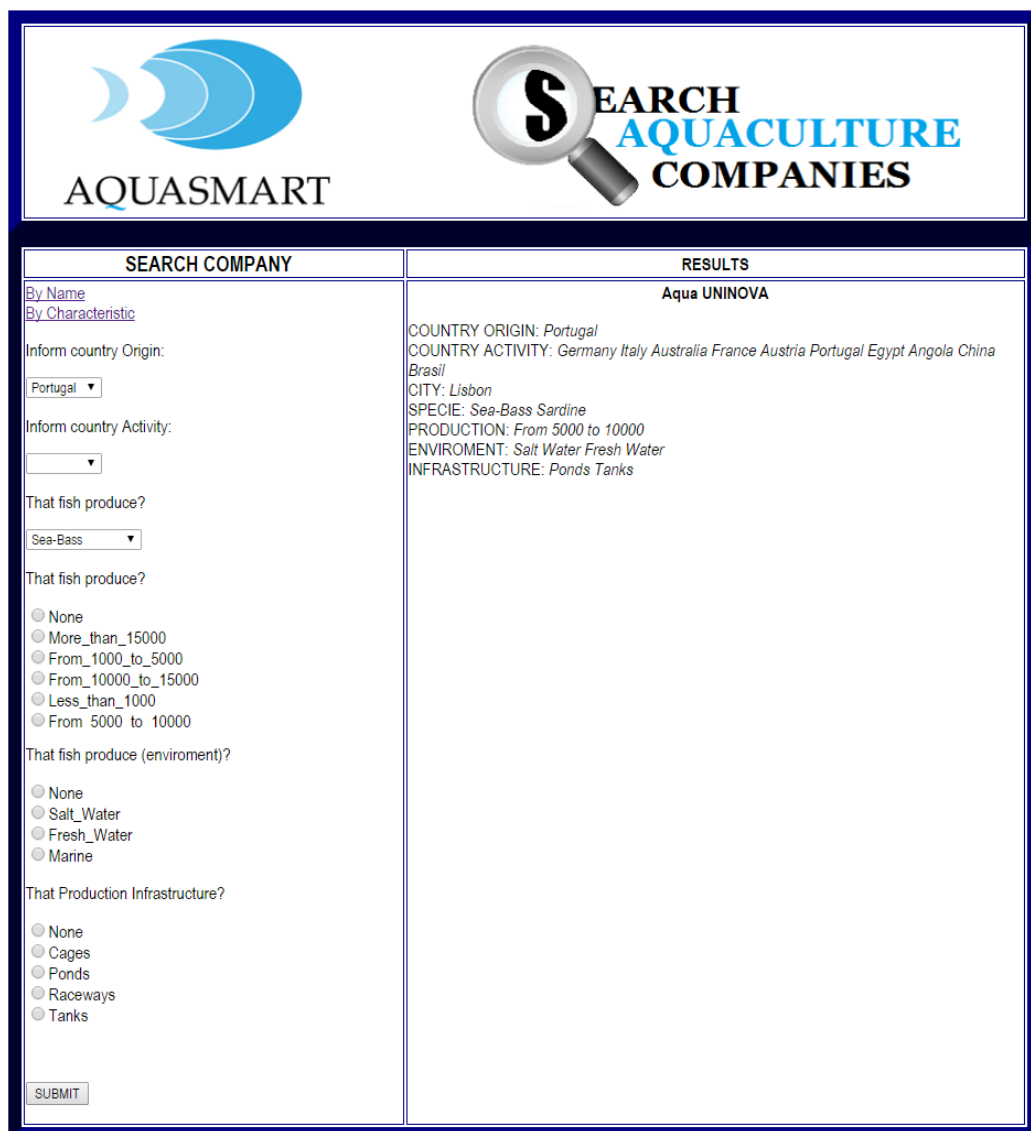
Figura 61 SE para diagnóstico de doenças

8.1.3 Sistema especialista para investigação de empresas de aquicultura

A fim de promover o desenvolvimento da aquicultura da UE, a Comissão Europeia elaborou orientações estratégicas, ajudando assim os Estados-Membros e as partes interessadas a vencer os desafios que o setor enfrenta. O setor da aquicultura da UE tem um

grande potencial de crescimento e pode contribuir para evitar a sobrepesca dos recursos marinhos.

Desta forma, o serviço “*Search by companies*” com pesquisa por características da empresa (Figura 62) e por nome das empresas (Figura 63) visa catalogar e pesquisar companhias que estão atuantes no mercado. Tal pesquisa ocorre por características específicas das companhias do setor da aquicultura, como espécies de peixes comercializadas, volume de produção, local de produção, dentre outras características. Assim, a *Search by companies* atua como interface do referencial semântico AquaSmart, no qual permite fazer inferências semânticas das companhias do setor da aquicultura.



SEARCH COMPANY	RESULTS
By Name By Characteristic Inform country Origin: Portugal ▼ Inform country Activity: ▼ That fish produce? Sea-Bass ▼ That fish produce? <input type="radio"/> None <input type="radio"/> More_than_15000 <input type="radio"/> From_1000_to_5000 <input type="radio"/> From_10000_to_15000 <input type="radio"/> Less_than_1000 <input type="radio"/> From_5000_to_10000 That fish produce (enviroment)? <input type="radio"/> None <input type="radio"/> Salt_Water <input type="radio"/> Fresh_Water <input type="radio"/> Marine That Production Infrastructure? <input type="radio"/> None <input type="radio"/> Cages <input type="radio"/> Ponds <input type="radio"/> Raceways <input type="radio"/> Tanks SUBMIT	Aqua UNINOVA COUNTRY ORIGIN: Portugal COUNTRY ACTIVITY: Germany Italy Australia France Austria Portugal Egypt Angola China Brasil CITY: Lisbon SPECIE: Sea-Bass Sardine PRODUCTION: From 5000 to 10000 ENVIROMENT: Salt Water Fresh Water INFRASTRUCTURE: Ponds Tanks

Figura 62 Sistema especialista para investigação de empresas de aquicultura - Características



 AQUASMART		 SEARCH AQUACULTURE COMPANIES
SEARCH COMPANY		RESULTS
By Name By Characteristic Inform company name: <input type="text" value="Andromeda_Group"/> <input type="button" value="SUBMIT"/>		Andromeda Group COUNTRY ORIGIN: <i>Greece</i> COUNTRY ACTIVITY: <i>Spain Greece</i> CITY: <i>Patras</i> SPECIE: <i>Sea-Bream Sea-Bass Meagre</i> PRODUCTION: <i>More than 15000</i> ENVIROMENT: <i>Marine</i> INFRASTRUCTURE: <i>Cages</i>

Figura 63 Sistema especialista para investigação de empresas de aquicultura - Nome